

Chapter 9

General Discussion

9

Objective of this thesis

The objective of this thesis was to develop and validate a comprehensive, generic, and short questionnaire to measure individual work performance (IWP). This objective was driven by the need to develop and evaluate interventions, procedures and strategies that can maintain, improve, or optimize IWP. An optimal IWP is crucial these days, due to increasing globalization of the economy and corresponding competitiveness between companies all over the world, the economic recession in many countries, and the growing need for sustainable employability. However, an existing lack of consensus in research and practice on how to define, conceptualize and measure IWP impeded valid measurement of the construct of IWP. In this final chapter, reflections on the main results and implications of this thesis are presented.

Overview of the main results

The first step towards a comprehensive, generic, and short measure of IWP was establishing a clear definition and conceptualization of IWP. In Chapter 2, conceptual frameworks of IWP were systematically reviewed from the occupational health, work and organizational psychology, and management and economics literature. The definition of IWP as “behaviors or actions that are relevant to the goals of the organization” [1] was chosen, because it is a broad definition of IWP, and thus, it is usable in multiple research fields, and applicable to a generic working population. Next, the underlying structure of IWP was determined. Numerous conceptual framework of IWP were integrated into one conceptual framework consisting of four broad and generic dimensions. The first dimension, task performance, can be defined as “the proficiency with which individuals perform central job tasks” [1]. The second dimension, contextual performance, can be defined as “behaviors that support the organizational, social and psychological environment in which the technical core must function” [2]. The third dimension, counterproductive work behavior, can be defined as “behavior that harms the well-being of the organization” [3]. Initially, adaptive performance was included as a fourth dimension, and was defined as “an employee’s ability to adapt to changes in a work system or work roles” [4]. Later on, however, the adaptive performance dimension was merged with the contextual performance dimension, leading to a conceptual framework consisting of three dimensions.

After the establishment of a clear definition and conceptualization of IWP, none of the existing questionnaires identified in the literature to measure (aspects of) IWP seemed adequate. For example, they did not capture the complete range of

individual work performance, did not include relevant content, or were not generically applicable. Therefore, it was decided to develop a new instrument that could overcome the limitations of existing instruments. In order to develop the Individual Work Performance Questionnaire (IWPQ), first, the operationalization of the IWPQ scales (task performance, contextual performance, and counterproductive work behavior) warranted attention. Indicators that could be used to measure the dimensions were identified from the scientific literature, existing questionnaires, and expert interviews (Chapter 3). From these, the most relevant indicators per dimension were selected by experts from different professional backgrounds. After a pilot-test, the IWPQ was subjected to a field-test in a large sample of blue, pink, and white collar workers (Chapter 4). After this, another improvement round was held in order to improve the targeting of the IWPQ. The questionnaire was tested again in a large sample of blue, pink, and white collar workers, and the final version of the questionnaire – the IWPQ 1.0 – was established (Chapter 5). The internal consistency and construct validity of the IWPQ 1.0 were good (Chapter 6). Conclusions about the responsiveness of the IWPQ cannot yet be drawn, and more research on this characteristic is necessary (Chapter 7). Also, the IWPQ 1.0 was cross-culturally adapted to the American-English language (Chapter 8). The main benefits of the IWPQ are that it measures all relevant dimensions of IWP, it is generically applicable, and it is short.

Methodological issue – generic applicability

A questionnaire to establish the effectiveness of interventions, procedures and strategies to maintain, improve, or optimize IWP, should be generically applicable, because such interventions often take place in varied settings. Thus, the same questionnaire needs to be suitable for different companies (e.g., so that results of an intervention can be compared across companies), and for a company with different types of employees (e.g., cleaners, office workers, managers). However, at the start of this thesis, it remained to be seen whether it was possible to develop such a questionnaire. Generic applicability of the questionnaire developed in this thesis was stimulated by conceptualizing and operationalizing IWP in a multi-disciplinary way, developing and evaluating the questionnaire using Rasch Analysis [5], and testing the questionnaire in a broad sample of blue, pink, and white collar workers.

In its current form, the IWPQ is generically applicable to workers from blue, pink, and white collar sectors. Although the responsiveness of the IWPQ deserves further attention, it is probable that the IWPQ is suitable for longitudinal

comparisons between the same group over time. This characteristic is most important, because the main purpose of the IWPQ is to establish the effectiveness of interventions, procedures and strategies to maintain, improve, or optimize IWP. In its current form, the IWPQ is less suitable for cross-sectional comparisons between different groups (e.g., comparing carpenters and dentists on IWP). Different cut-off points should be used when interpreting scores for workers from different occupational sectors, because the IWPQ includes a few items with Differential Item Functioning (DIF; see Chapters 4 and 5). Items with DIF were allowed in the IWPQ, because in order to detect changes in IWP over time, adequate targeting is more important than DIF-free items. Thus, the IWPQ is generically applicable, but we should be cautious in comparing different occupational groups on IWP.

The IWPQ scales were developed and validated in three broad occupational sectors, namely blue, pink, and white collar workers. As a consequence, we could not examine the reliability and validity of the IWPQ in every specific occupation. Although generic questionnaires pose considerable advantages in research, when examining IWP in a specific job, a job-specific questionnaire may be preferred to capture all the relevant aspects of that job. When such a questionnaire is not available, job-specific questions could be added to the IWPQ. An advantage of doing this is that all relevant dimensions of IWP are included (which often are not in existing questionnaires, as examined in Chapter 3).

Methodological issue – use of self-report

The IWPQ was developed as a self-report questionnaire. Self-reports can be accompanied by several biases, such as *recall bias*, *social desirability bias*, *leniency effects*. Due to recall bias, systematic error may be introduced in the answers by inaccuracy or incompleteness of people's recollections of their past behaviors at work [6]. Also, a questionnaire on IWP may elicit socially desirable answers. That is, people tend to respond to an item in a certain way, because they think it is the socially acceptable answer, rather than their true answer [7]. A closely related bias is the leniency effect, that is, people are naturally motivated to present themselves in a favorable, positive light [6]. As a result of social desirability and leniency effects, self-ratings of performance may lead to a higher score than in reality. For example, Van der Heijden and Nijhof [6] found that self-ratings of performance are generally one half to one standard deviation higher than ratings by peers or managers.

In the IWPQ, several precautions were taken to minimize influences of self-report biases. For example, effects of social desirability and leniency were minimized

by leaving out the questionnaire title and scale names when administering the questionnaire, so that negative connotations related to IWP are avoided, and by reassuring participants that their answers are anonymous and treated confidentially. In addition, the center of the answer category scale was shifted, so that the average point was not in the middle but rather to the left of the scale for task and contextual performance, and to the right of the scale for counterproductive work behavior. Doing this will prevent possible floor and ceiling effects that may result from social desirability and leniency biases [8].

Despite possible biases accompanying self-report, the IWPQ was developed as a self-report questionnaire for several reasons. First, in many occupations, objective measures of performance are not easily obtainable [9]. Especially for knowledge work or high complexity jobs, direct measures of countable behaviors or outcomes, such as production quantity or number of errors made, are almost impossible. Second, self-reports have practical advantages such as ease of collection, issues of confidentiality, and less problems with missing data, when compared to objective measures, peer or managerial ratings [10]. Finally, peer or managerial ratings of performance can also be accompanied by several biases, such as *under-sampling bias*, *halo effects* and *leniency effects*. In comparison to self-ratings, ratings by supervisors are based on a much smaller amount of information, leading to the so-called effect of *under-sampling*. The person who is doing the job possesses the greatest familiarity with the job and their own behavior at work, and because of that, is an appropriate person to fill in the questionnaire [6]. The notion that employees have more opportunity to observe their own behaviors than peers or managers do may be especially true for counterproductive behaviors, because most of these behaviors are intended to be private and, hence, unobservable [11]. A recent study found that self-raters actually reported engaging in more counterproductive behaviors than other-raters reported them engaging in [12]. Self-report methodology is consistently used throughout CWB research, and given limitations in other methods, several studies have concluded that it is the most appropriate method [e.g., 13, 14]. Another bias accompanying peer or supervisor ratings of performance is the *halo effect*, in which the peer's or supervisor's general impression of the employee (for example, liking or disliking the employee) influences the evaluation. As a result, peers and supervisors score the different dimensions of IWP are more similar (i.e., they answer more consistently), and inter-correlations between the dimensions are overestimated [6, 11, 15]. In addition, not only self-ratings, but also peer and supervisor ratings are influenced by *leniency effects*.

Although research has shown that this effect is stronger for self-ratings, it is also present in peer and supervisor ratings [16, 17]. Thus, peers and supervisors tend to see others in a favorable, positive light [6].

Methodological issue – lack of a golden standard

The criterion validity of the IWPQ scales could not be examined, because there was no golden standard available. This makes it difficult to prove that a new measure is indeed measuring what is intended [8]. A perfect golden standard may be an objective measure of IWP, but in many occupations these are hard to obtain [9]. Especially for knowledge work or high complexity jobs, direct measures of countable behaviors or outcomes, such as production quantity or number of errors made, are almost impossible. Other possible golden standards may be peer or managerial ratings. However, peer or supervisor ratings are accompanied by several biases, as described above, and therefore, cannot be considered golden standards of IWP. The solution lies in examining the construct validity of the questionnaire, as was done in the current thesis. Although construct validity is often considered to be less powerful than criterion validation, with strong theories and specific and challenging expectations, it is possible to acquire substantial evidence that the measurement instrument is measuring what it purports to measure [18].

Strengths

IWP is an issue that is researched in multiple research fields, and its conceptualization and operationalization must ideally be considered in a multi-disciplinary way. A strength of the current thesis is that multiple research fields and stakeholders were involved throughout the development process of the IWPQ. First, literature from multiple research fields was used in order to establish a definition and conceptual framework of IWP. Second, the literature, existing questionnaires, and experts from multiple research fields were consulted in order to construct the questionnaire. Third, workers from different occupational sectors were included in the field-testing phase, resulting in a generically applicable questionnaire.

A second strength of the current thesis was that a thorough development and improvement process was applied before the final IWPQ was reached. Often, a major problem with many studies is that insufficient time is allowed for proper field-testing, further adaptation and re-evaluation before the final instrument is used in research and/or practice [18, 19]. As described in more detail in the paragraphs above, in the current thesis, a clear definition and conceptual framework of IWP

were established before developing the questionnaire, consensus was reached on the operationalization of the scales, and a pilot-test, field-test, and improvement round were held. Subsequently, the internal consistency and validity of the final IWPQ scales were tested. This has resulted in a reliable and valid questionnaire that measures IWP in a short, comprehensive and generic way, and that is ready to be used in research and practice.

Another strength of this thesis is that it is the first to develop and evaluate an IWP questionnaire using Rasch analysis [5]. This offered unique insights into the IWPQ scale characteristics. Rasch analysis ensures that key measurement assumptions, such as appropriate category ordering, local independence, and differential item functioning, are tested. In addition, Rasch analysis has particular value in the development of new questionnaires, specifically in guiding item reduction [20]. Furthermore, it ensured that items were suitable for all occupational sectors, and gave insight into whether the items were well distributed over the whole range of the scale (targeting). When there is optimal targeting, one can reliably measure persons at all levels of ability, and discriminate between persons at various ranges on the scale (e.g., discriminate amongst workers with low performance, as well as amongst workers with high performance). When improving the targeting of the IWPQ, Rasch analysis guided the removal of misfitting items and the addition of new items that improved targeting. A final benefit of Rasch analysis is that it provides a statistically proven interval level scale, instead of an ordinal level scale that is formed by the raw scores [21]. This is useful when one wants to measure changes in IWP over time. In an ordinal scale, higher scores indicate higher performance, but the relative distances between the scores are meaningless [22]. For example, it is unclear whether a person whose performance increased from 2 to 3, has made the same amount of improvement as a person whose performance increased from 3 to 4. On the contrary, an interval scale allows straightforward interpretation of the distances between scores, so that a change in score from 2 to 3 is equivalent to a change from 3 to 4. Thus, a benefit of the IWPQ is that its change scores can be estimated and interpreted more accurately than other IWP questionnaires using Rasch analysis, because it has a statistically proven interval scale.

Limitations

A limitation of the present thesis is that the responsiveness of the IWPQ remains unknown. So far, the IWPQ was used in an intervention study that stimulated

physical activity and relaxation amongst office workers, in order to improve their need for recovery and work-related outcomes (such as IWP). In this intervention study, statistically significant changes in tasks performance and contextual performance over time were detected with the IWPQ [23]. However, no firm conclusions could be drawn about the responsiveness of the IWPQ based on the results of that study. Several reasons may have accounted for this, such as a study population with high baseline scores, small changes on many constructs in the study, and unclear guidelines for interpreting responsiveness. Nevertheless, there is good reason to believe that the IWPQ has good responsiveness, because Rasch analysis ensured that items with a high discrimination parameter were included in the IWPQ (see Chapters 4 and 5), the IWPQ showed very good discriminative validity (see Chapter 6), and significant changes in IWP over time were detected in an intervention study [23].

A second limitation is that the IWPQ task performance scale has ceiling effects, and the CWB scale has floor effects. Despite an improvement round to improve the targeting of the scales (see Chapter 5), workers still score relatively high on task performance, and relatively low on CWB. As a consequence, the IWPQ is less able to discriminate workers with high task performance, and less able to discriminate workers with low counterproductive performance. Especially in longitudinal studies, this could be problematic, because workers who at baseline score high on task performance, or low on CWB, cannot show any further improvement (thus, even more task performance, or even less CWB). As previous examination of the IWPQ using Rasch analysis showed that the items of the IWPQ are relatively well-distributed over the scales (see Chapter 5), the ceiling and floor effects are not likely to be a shortcoming of the measurement instrument, but rather, a true characteristic of the population (an actual high/low occurrence of these behaviors in the workplace). Furthermore, social desirability and leniency biases may drive workers to present themselves in a favorable light. When administering the IWPQ, it is therefore important to leave out the questionnaire title and scale names to avoid negative connotations related to IWP, and to guarantee participants' anonymity.

A final limitation is that the IWPQ is unsuitable for individual level use. For individual level use, a minimum reliability value of 0.90, and preferably 0.95, is required. For group level use, a minimum reliability value of 0.70 is required [18, 24, 25]. As the reliability of the IWPQ scales varies from 0.74 for the CWB scale to 0.85 for the contextual performance scale, the IWPQ can currently only be used for cross-

sectional comparisons between, and longitudinal comparisons within, groups. This means that the IWPQ should not be used for comparisons at the individual level, nor for assessments, evaluations, and selection processes at the individual level.

Recommendations for future research on the IWPQ

As evidenced by the current thesis, the development and validation of a questionnaire can easily take years, and even then, it is never quite done. Strictly speaking, one can never state that an instrument is valid, only that it provides valid scores in the specific situation in which it has been tested [18]. As construct validation is an ongoing process, more research should be conducted to create a strong web of evidence to support the validity of the IWPQ. The relationship of the IWPQ scales with other constructs, such as work engagement, job satisfaction, and health, should be examined in settings and populations other than was done in the current thesis. Also, despite the limitations of peer and supervisor ratings, and objective measures of performance, it would be interesting to compare these with self-ratings on the IWPQ scales for validation purposes.

In addition, more research is necessary to draw conclusions about the responsiveness of the IPWQ. Preferably, its responsiveness is examined in intervention studies that directly target IWP, in populations that have low baseline levels on the constructs under examination. Suggestions for such populations could be workers with work-related musculoskeletal health problems, mental health problems, and/or low job satisfaction. An intervention study, which is directly aimed at improving IWP, could obtain greater changes in these populations, making it easier to detect changes in IWP and related constructs. Suggestions for such a study could be an intervention focusing on managerial style, technological improvements at work, and/or job skills training.

For both research and practice, it is important that the scores on the IWPQ scales can be easily interpreted. In Chapter 5, an interpretation is given of *single scores* from “very low” to “very high” performance. However, it remains a question for future research how *change scores* should be interpreted. To do this, the smallest detectable change (SDC) and the minimally important change (MIC) should be determined. The SDC represents the change that can be detected beyond measurement error, and can be determined using the standard error of measurement. To obtain the standard measurement error of the IWPQ scales, test-retest reliability of the scales should be determined in a stable population over a short time interval (e.g., a couple of days). The MIC represents the smallest change

which individuals perceive as important, and should ideally be determined with someone who can assign practical meaning to a change, for example the employees themselves, their colleagues, or their supervisor(s). The MIC could, for example, be expressed in a number of points change, or a percentage change, from the baseline score on the IWPQ scales.

As of yet, the IWPQ is only suitable for group use. Future research could examine whether the IWPQ scales can be adapted for individual use. A strategy could be to add more questions to the scales in order to obtain higher reliability (a minimum value of 0.90 is required for individual use [18, 25]). It remains to be seen whether such questions can be generic across occupations, or whether company- or job-specific questions are more appropriate. When the reliability of the IWPQ scales is high enough for individual use, the IWPQ can be used for cross-sectional comparisons between, and longitudinal comparisons within, individuals. Participants' anonymity and confidentiality of answers should still be guaranteed, so that employees do not have to worry about their answers being seen by their supervisor(s) or colleague(s).

At a group level, tailored feedback could be given to managers or employers who want to monitor, benchmark and improve their own team, department or company, respectively. They could be given an overview of their team's, department's, or company's results, benchmarked against mean scores in similar groups. Future research should determine which advice can be given to managers and employers on how to improve their team, department or company scores on the IWP dimensions.

Finally, IWP may be influenced by the Zeitgeist. In other words, it may not be a stable construct over time. Changes in work, society, and the economy can affect what exactly constitutes performance at work. For example, whereas task performance was considered the main component of IWP in the past, contextual performance (and its variants) have gained more and more currency since the 1960s. Also, adaptive performance behaviors have gained ground into IWP theories and questionnaires, either as a distinct dimension or as part of the contextual performance dimension. Adaptive performance items that have come up in questionnaires because workers need to be increasingly adaptive, versatile, and tolerant, are for example: *"Do you periodically update your skills to accomplish the work or projects you are assigned?"* or *"Do you make friends with people from different countries?"* [26]. In one or two decades from now, the content of the dimensions may have changed, other dimensions may have arisen (e.g.,

environmental sustainability), and/or some existing dimensions may become obsolete. It is therefore important to constantly evaluate and scrutinize what the concept of IWP means, and whether we still measure what we want to measure.

Towards an integration of research fields

Multiple research fields, such as occupational health, work and organizational psychology, and management and economics, are interested in IWP. All have developed their own way of approaching IWP, ways that have influenced why and how measurement instruments are constructed in the first place, and the manner in which they are interpreted. The array of available questionnaires to measure (aspects of) IWP is daunting when searching the literature. Even within one research field, it seems that every study has used a different approach to measure (aspects of) IWP. This proliferation of instruments impedes the interpretation of study results, since findings cannot be compared with each other [8].

In order to advance research on IWP, an integration of the different research fields is proposed. The conceptualization and operationalization of IWP proposed in this thesis provides a good starting point for that, as they were based on all these research fields. We recommend researchers in all research fields to adopt the definition and conceptual framework of IWP proposed in this thesis. Secondly, the IWPQ may be used as the standard instrument to measure IWP in a comprehensive, generic, and short way. Below, it is described what an integration could add to each research field.

In the field of occupational health, the terms absenteeism, presenteeism, and productivity are often used interchangeably with IWP. Consensus on their definitions, and what sets them apart from each other, would bring clarity to the field. Current preventive intervention programs in occupational health, applied to a generic working population, focus on absenteeism and presenteeism as outcome measures, while most of the employees in the generic working population have low rates of absenteeism and presenteeism to start with. As a result, absenteeism and presenteeism instruments suffer from floor and/or ceiling effects, making it very difficult to find further improvements due to interventions. Using the IWPQ as the standard instrument to measure IWP, the field of occupational health will benefit from looking at a wider range of performance behaviors at work. This will enable researchers to show that occupational health interventions are not only suitable for maintaining, improving and optimizing IWP of employees that are frequently absent, or have reduced performance due to health complaints, but also of the larger, mostly

healthy, population of employees. A finding to support the additional value of looking at IWP, is that supervisor-rated performance at work is not that strongly related to the number of days that one is absent from work due to health complaints [27, 28]. This indicates that IWP and absenteeism are two distinct concepts, and both provide unique and valuable information. Furthermore, the field of occupational health could learn from the field of management and economics, by including company-level performance indicators in addition to individual-level indicators such as absenteeism, presenteeism and IWP.

In the field of work and organizational psychology, a lot of research has been done on how to define and measure IWP. A lot of work in the current thesis was drawn from work and organizational psychology. For example, the definition that was adopted in the present thesis originated from work and organizational psychologist Campbell [1]. The long research tradition of work and organizational psychology in the area of IWP has resulted in a lot of different labels and measures for the concepts of task performance, contextual performance, and counterproductive work behavior. The IWPPQ could function as the standard measurement instrument for (the dimensions of) IWP, which would increase generalizability and comparability of findings, and avoid problems of existing measurement instruments from work and organizational psychology, such as a lack of content validity of the scales, and overlapping items between scales (leading to unrealistically high correlations between scales). In addition, work and organizational psychology could learn from occupational health, and management and economics, for example by including measures of absenteeism, and indicators of company-level performance, such as employee turnover, customer satisfaction, and financial performance.

In the field of management and economics, performance measurement mostly focuses on the company level, using key performance indicators such as employee turnover, customer satisfaction, and financial performance. This research field could learn from occupational health, and work and organizational psychology, to include IWP, and individual level outcomes such as absenteeism, employee health and well-being, and work engagement, in addition to company level outcomes. This would give insight into the relationship between IWP and company level outcomes – a relationship where there is currently little information on. In addition, including the IWPPQ as an outcome measure can determine the effectiveness of company-level interventions, methods and strategies on IWP. This can also give insight into the

possibility that certain strategies might work for some groups of employees but not for others, under which conditions, and why.

Applications for practice

As stated previously, the main purpose of the IWPQ is research-oriented, namely, to establish the effectiveness of interventions, procedures and strategies to maintain, improve, or optimize IWP. Nevertheless, IWP is not only an important outcome measure in research, but also in practice. Words such as performance goals, performance feedback, and performance reviews, are familiar terms in many workplaces. The knowledge gained in the current thesis can be applied in practice in multiple ways. First of all, although the IWPQ is currently not suitable for individual assessments, evaluations, and selection processes, professionals dealing with IWP in the workplace (e.g., managers, human resource managers, and occupational health physicians) can keep in mind the three dimensions of IWP, when observing, evaluating, or improving IWP of employees. Both managers and employees can keep in mind the three dimensions of IWP when going into a performance review and formulating performance goals. For example, they should not only look at whether central job tasks are performed, but also at taking on extra tasks, keeping job knowledge and skills up-to-date, or showing excessive negativity in the workplace. Also, managers should be aware that these dimensions are not necessarily strongly related (i.e., an employee can perform contextual behaviors and counterproductive work behaviors simultaneously).

Second, the knowledge gained in the current thesis could benefit HRM and occupational physicians in companies. Core tasks of HRM include hiring and recruiting new personnel, assessing personnel, and providing training and development programs for personnel. HRM could use the IWPQ to identify groups of employees that perform well on the job (e.g., groups with certain personality traits, or skills), so that this knowledge can be used to hire and recruit new personnel. HRM could use the conceptual framework of IWP proposed in the current thesis (i.e., they should consider task performance, contextual performance, and counterproductive work behavior) when assessing personnel. Furthermore, HRM could use the IWPQ to determine which groups of employees need to improve on which aspect(s) of IWP, and subsequently, which training and development program(s) would be most suitable for those groups. Core tasks of occupational health and safety departments include providing a safe and healthy work environment, guiding sick employees in order to promote return to work, and

monitoring health of employees (e.g., health risk appraisals). Occupational physicians could use the IWPQ to monitor IWP of groups of employees reintegrating to the job after a period of sick leave.

Third, the knowledge gained in the current thesis is interesting for companies. They can include the IWPQ as a key performance indicator, in addition to existing ones such as employee turnover, customer satisfaction, financial performance. This way, the IWPQ can be used to monitor IWP. Also, by combining the IWPQ with other company-level measures, the relationship between IWP and company productivity and competitive ability can be untangled, and effective ways of improving company productivity and competitive ability through improving IWP can be identified.

Trends in work

A large part of the application of the IWPQ lies in the ability of research to answer questions that arise in practice. Current trends in work, such as those described in the General Introduction (globalization, economic recession, and sustainable employability), pose challenges to society that research can help overcome. In order to overcome these challenges, collaboration between research and practice is necessary. Knowledge gained on the predictors and effects of IWP in research can be used to develop interventions, procedures and strategies that are effective in maintaining, improving and optimizing IWP. Such interventions should be developed together with practice (e.g., employees, managers, HRM). Ultimately, this will improve the competitive ability of companies, optimize IWP in times of economic recession, and maintain employability up to an older age. Below, the questions that arise with each trend in work, and the role that research could play in answering them, are described.

Due to globalization, competition between companies from all over the world increases. Therefore, it is essential for companies to improve IWP of their employees, and thereby, their productivity and competitive ability. The IWPQ could be used to answer questions from practice regarding the effects of globalization. For example, globalization is heavily linked with advances in technology [29]. A question that might arise here, is if it is beneficial for IWP if employees are available on their smartphones and laptops 24/7? Does technology aid or hurt employee work engagement, work-home balance, and need for recovery after a workday, and how does this in turn affect IWP? Finally, the IWPQ can be used to gain knowledge on the predictors of IWP, and discover effective ways to improve IWP. For example, how

can technology aid employees, and when so (e.g., faster processes, higher quality) and when not (e.g., less personal communication, dependency on technology)?

Due to the current economic recession, it is important for companies to increase IWP of employees in order to stay afloat. Often, companies see their profits decrease, and in order to survive, they have to cut costs. This is often accomplished by employee reductions, reorganizations, or outsourcing work to cheaper markets. Also, employee salaries can be frozen, and bonuses downsized or eliminated. Such measures are often stressful for employees, lead to increased job insecurity, and subsequently, may have negative effects on IWP. On the other hand, employees might tend to increase their IWP in order to increase their chance to survive possible reorganizations. The IWPQ can be used to monitor the effects of the economic recession on IWP at the group level. It should be noted that, in times of economic recession, it is questionable whether employees will fill out the questionnaire honestly, because they are afraid there may be consequences if they perform poorly (e.g., job loss). In addition to monitoring IWP, research can help answer the question of how companies can increase IWP of employees, despite the economic recession and its possible negative influences on, for example, employee morale, stress, and job security. New Ways of Working is a recent and popular strategy to boost IWP, by giving employees more freedom and flexibility, while at the same time lowering company costs. The IWPQ could be used to determine how New Ways of Working influences IWP. Also, supplemented with additional measures, it could be used to examine which parts of New Ways of Working are especially effective in increasing IWP, and for whom (e.g., effectiveness dependent on age or personality?).

Third, it is increasingly important to promote sustainable employability of older workers in the workforce. The retirement age of workers in Europe will be increased in coming years [30], which means that employees have to work at the same – or even higher – level of IWP, despite possible limitations caused by an older age, such as reduced health. Several studies have found that both negative and positive stereotypes are associated with older workers [e.g., 31-33]. On the one hand, older workers are believed to be less willing to adapt to change, less motivated to learn, and less productive than younger workers. On the other hand, they are perceived to be more reliable, more loyal to the organization, and more experienced and knowledgeable than younger workers. The IWPQ can be used to monitor IWP with age. By including additional measures, predictors of older workers' IWP can be identified. To what extent do health and vitality impact IWP of older workers? Do older employees have difficulty to adapt to and keep up with advances in new

technology, and how does this affect their IWP? Knowledge on the predictors of IWP, and how we can influence them, can tell us how we can maintain IWP up to an older age. What can employees do to stay employable? How can employers facilitate them in staying employable? Are interventions, for example, focusing on a healthy lifestyle and vitality, organizational processes, or New Ways of Working, effective in maintaining, improving and optimizing IWP of older workers? How can technology be used to maintain, improve and optimize IWP of older workers?

A multi-disciplinary challenge

Maintaining, improving and optimizing IWP is a multi-disciplinary challenge, and we have the best chance of competing in a globalizing environment, overcoming the economic recession, and promoting sustainable employability when we approach this challenge in a multi-disciplinary way. With increased integration between research fields, and standardized measurement of IWP, predictors of IWP can be established even more precisely and comprehensively. Different research fields should draw on each other when developing interventions, procedures and strategies to maintain, improve, or optimize IWP, because research has shown that interventions with a multi-disciplinary approach are often most successful [e.g., 34, 35]. Finally, interventions, procedures and strategies should not be developed only by researchers, but in collaboration with practice (e.g., employees, managers, HRM). Cross-over of information between research fields and practice creates 'out of the box' thinking and can result in solutions that normally would not have been thought of.

Conclusion

The objective of this thesis – to develop and validate a comprehensive, generic, and short questionnaire to measure IWP – was achieved. The Individual Work Performance Questionnaire (IWPQ) is based on a conceptual framework consisting of three dimensions, namely, task performance, contextual performance, and counterproductive work behavior. The questionnaire is suitable for generic use (workers in all types of occupations, and workers with and without health problems) and is short with only 18 items. The internal consistency of the IWPQ was good and the construct validity was acceptable. Future research is necessary to determine the responsiveness of the IWPQ. A Dutch and American-English version of the IWPQ are available.

The current thesis contributes towards consensus on the definition and conceptualization of IWP, and hopefully, provides a push towards increased integration between research fields interested in IWP. Furthermore, it provides a reliable and valid instrument to measure IWP in a comprehensive, generic, and short way. Standardized measurement of IWP allows comparability and generalizability of findings, and increases knowledge on predictors and effects of IWP. This knowledge can be used to develop interventions, procedures and strategies to maintain, improve, or optimize IWP, and subsequently, evaluate their effectiveness. So, if you really want to know whether the promise of a financial bonus, being available on your mobile 24/7, or feeling happy, increases performance at work, as the news headings and articles in the General Introduction pose, use the IWPQ! As Lord Kelvin (1883) said: "*Measurement is knowledge*".

References

1. Campbell JP. Modeling the performance prediction problem in industrial and organizational psychology. In: Dunnette MD, Hough LM, editors. *Handbook of industrial and organizational psychology, Vol.1* (2nd ed.). Palo Alto, CA, US: Consulting Psychologists Press; 1990. p. 687-732.
2. Borman WC, Motowidlo SJ. Expanding the criterion domain to include elements of contextual performance. In: Schmitt N, Borman WC, editors. *Personnel Selection in Organizations*. San Francisco, CA: Jossey Bass; 1993. p. 71-98.
3. Rotundo M, Sackett PR. The relative importance of task, citizenship, and counterproductive performance to global ratings of performance: A policy-capturing approach. *Journal of Applied Psychology*. 2002;87(1):66-80.
4. Griffin MA, Neal A, Parker SK. A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal*. 2007;50:327-47.
5. Rasch G. *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press; 1960.
6. Van der Heijden BIJM, Nijhof AHJ. The value of subjectivity: Problems and prospects for 36-degree appraisal systems. *The International Journal of Human Resource Management*. 2004;15(3):493-511.
7. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*. 2003;88(5):879-903.
8. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development*, 4th ed. Oxford University Press; 2008.
9. Jaramillo F, Carrillat FA, Locander WB. A meta-analytic comparison of managerial ratings and self-evaluations. *Journal of Personal Selling & Sales Management*. 2005;XXV(4):315-28.
10. Schoorman DF, Mayer RC. The value of common perspectives in self-reported appraisals: You get what you ask for. *Organizational Research Methods*. 2008;11(1):148-59.
11. Dalal RS. A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*. 2005;90:1241-55.

12. Berry CM, Carpenter NC, Barratt CL. Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*. 2012;97(3):613-36.
13. Fox S, Spector PE. A model of work frustration-aggression. *Journal of Organizational Behavior*. 1999;20(6):915-31.
14. Coyne I, Gentile D, Born M, Ersoy NC, Vakola M. The relationship between productive and counterproductive work behaviour across four European countries. *European Journal of Work and Organizational Psychology*. 2013;22(4):377-89.
15. Viswesvaran C, Ones DS. Perspectives on models of job performance. *International Journal of Selection and Assessment*. 2000;8(4):216-26.
16. Holzbach RL. Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology*. 1978;63(5):579-88.
17. Thornton GCI. Psychometric properties of self-appraisals of job performance. *Personnel Psychology*. 1980;33:263-71.
18. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine*. Cambridge University Press; 2011.
19. Sudman S, Bradburn NM. *Asking questions - A practical guide to questionnaire design*. 1st ed. Sudman S and Bradburn NM, editors. San Francisco, USA: Jossey-Bass Publishers; 1982.
20. Tennant A, McKenna SP, Hagell P. Application of rasch analysis in the development and application of quality of life instruments. *Value in Health*. 2004;7:S22-6.
21. Prieto L, Alonso J, Lamarca R. Classical test theory versus rasch analysis for quality of life questionnaire reduction. *Health Quality of Life Outcomes*. 2003;1(27).
22. Embretson SE, Reise SP. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers; 2000.
23. Coffeng JK, Hendriksen IJM, Duijts SF, Proper KI, Van Mechelen W, Boot CRL. Effectiveness of a combined social and physical environmental intervention on work-related outcomes in office employees, submitted.
24. Tennant A, Conaghan PG. The rasch measurement model in rheumatology: What is it and why use it? when should it be applied, and what should one look for in a rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research)*. 2007 12/15;57(8):1358-62.

25. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
26. Pulakos ED, Schmitt N, Dorsey DW, Arad S, Hedge JW, Borman WC. Predicting adaptive performance: Further tests of a model of adaptability. *Hum Perform*. 2002;15(4):299-324.
27. Bycio P. Job performance and absenteeism: A review and meta-analysis. *Human Relations*. 1992;45(2):193-220.
28. Morrow PC, McElroy JC, Laczniak KS, Fenton JB. Using absenteeism and performance to predict employee turnover: Early detection through company records. *Journal of vocational behavior*. 1999;55:358-74.
29. Joshi RM. *International business*. Oxford University Press, New Delhi and New York; 2009.
30. European Commission. *EUROPE 2020: A strategy for smart, sustainable and inclusive growth*. Brussels: European Commission; 2010.
31. McGregor J, Gray L. Stereotypes and older workers: The New Zealand experience. *Social Policy Journal of New Zealand*. 2002(18):163-77.
32. Maurer TJ, Wrenn KA, Weiss EM. Toward understanding and managing stereotypical beliefs about older workers' ability and desire for learning and development. *Research in Personnel and Human Resources Management*. 2003;22:253-85.
33. Bal AC, Reiss AEB, Rudolph CW, Baltes BB. Examining positive and negative perceptions of older workers: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2011;66B(6):687-98.
34. Cancelliere C, Cassidy JD, Ammendolia C, Côté P. Are workplace health promotion programs effective at improving presenteeism in workers? A systematic review and best evidence synthesis of the literature. *BMC Public Health*. 2011;11:395.
35. Kahn-Marshall JL, Gallant MP. Making healthy behaviors the easy choice for employees: A review of the literature on environmental and policy changes in worksite health promotion. *Health Education & Behavior*. 2013;39:752-76.