Taylor & Francis
Taylor & Francis Group

The distinction between work pace and working hours in the measurement of quantitative demands at work

TAGE S. KRISTENSEN[★], JAKOB B. BJORNER,
KARL B. CHRISTENSEN and VILHELM BORG

National Institute of Occupational Health, Lerso Parkalle 105, DK-2100
Copenhagen O, Denmark

During recent years many researchers have criticized the widely used scales on psychological job demands. For instance, they comment that in most cases different types of demand seem to be mixed in one measure. In this paper we analyse the scale on quantitative job demands in the recently developed Copenhagen Psychosocial Questionnaire (COPSOQ), with special emphasis on Differential Item Functioning (DIF). DIF refers to basic differences between groups of respondents, which may affect how they respond to questionnaire items. The data material for our study comprised a representative sample of Danish employees. The respondents were categorized into 32 specific jobs according to the International Standard Classification of Occupations (ISCO 1968). We analysed DIF with respect to the respondents' jobs with logistic regression analyses. These analyses showed that the items used in the original demand scale functioned very differently for different jobs in the population. The conclusion is that scales on quantitative demands are very sensitive to the choice of specific items. If many items on fast work pace and tempo are included in a scale, a number of blue-collar jobs will be identified as high-demand jobs. If, on the other hand, many questions on long working hours and overtime are included, the use of the scale will result in an entirely different picture. This issue has so far received little attention in occupational health psychology. The results have wide theoretical and methodological implications for research on quantitative job demands.

1. Introduction

Psychological job demands probably constitute the most important single factor in the field of psychosocial work environment research. This seems to be the case whether we look at the theories of person-environment fit, the demand-control model, the burnout research, or the model of effort-reward imbalance. All theories or models seem to focus on the balance—or lack of balance—between demands at work and something else, be it personal resources, decision latitude, social support, coping strategies, or rewards. While the literature on job demands is abundant, there have been very few contributions focusing on *theoretical* and *methodological* issues relating to the concept of psychosocial job demands. This

---

[★]Author for correspondence. e-mail: tsk@ami.dk

seems to be one of the major paradoxes of recent research in psychosocial factors at work. In the present paper we will focus on the concept of *quantitative demands* and we will base our discussion of validity and measurement on an analysis of *differential item functioning* (DIF) in a newly-developed scale on quantitative demands at work.

1.1. *The concept of quantitative demands*

In the international literature on psychosocial factors at work the discussion about demands at work has been going on for some years. One of the most striking features of the whole literature on job strain/job demands is the almost complete lack of clear conceptual definitions. In spite of the fact that the job strain model has been the dominating model in occupational health psychology for about 25 years, the basic question about a clear definition remains unanswered. In the now classical paper of 1979, Karasek loosely defines demands as 'work load demands, conflicts or other stressors which place the individual in a motivated or energized state of "stress"' (p. 287) and as 'the psychological stressors involved in accomplishing the work load, stressors related to unexpected tasks, and stressors of job-related personal conflict' (p. 291). In the book by Karasek and Theorell (1990), job demands are briefly defined as 'how hard you work' (p. 63), but the authors also add the following prudent remark: 'Indeed, psychological demands on the job remain difficult to conceptualize and measure because of the diversity of subcomponents and because of some theoretical problems that are as yet unresolved' (Karasek & Theorell, 1990, p. 63).

During the last 5–6 years a number of researchers have expressed concern that the demand concept and the measurement of demands at work need to be refined and improved (de Jonge & Kompier, 1997; Hallqvist, Diderichsen, Theorell, Reuterwall, & Ahlbom, 1998; Johnson, Stewart, Hall, Fredlund & Theorell, 1996; Steenland, Johnson, & Nowlin, 1997; Theorell *et al.*, 1998; van der Doef & Maes, 1999). Among the points raised by these and other writers are that the demand concept is poorly defined, that several types of demands seem to be mixed in one measure, that the same questionnaire items may have different meanings for different respondents, and that some of the most widely used scales have low reliability and validity. Most of these comments have been made in relation to research based on the demand-control model, but the general theoretical and methodological issues are relevant for other models and measures as well. A number of widely used scales for measuring (quantitative) job demands are shown in Figure 1.

The basic definition of *validity* is that *we measure what we want to measure*. This immediately makes it clear that it is meaningless to discuss the validity of a scale if it is unclear what the scale is intended to measure. A number of authors have commented on this problem in recent papers. A good example is the comments by Steenland *et al.* (1997) on a negative study of job strain and heart disease: 'The variable for job demands may be measuring something else than what was originally intended—demands for fast-paced performance. In our data, psychological demands are positively correlated with both income and education. Occupations with high demand scores may be those which require more challenging and mentally active work—components of the work process that are more health enhancing than otherwise' (Steenland *et al.*, 1997, p. 259). In that study a job exposure classification based on the five items of the Job Content Questionnaire (JCQ) was used. The study showed that high demands were associated with *lower* risk of heart disease. The comment by Steenland *et al.* (1997) is interesting because it states that the JCQ may be measuring something other (challenging work) than it was originally intended to measure. According to Steenland *et al.* (1997) the scale was intended to measure fast-paced work, but this statement is hardly correct since it assumes that a strict definition exists in the literature.

A. Alfredson, Karasek & Theorell, 1982:
   'Unemployed at least once during the past 5 years'
   'Working at least 10 h every day'
   'At least one third of total work income from piece wage'
   'Hectic work'
   'Continuously changing day and night work schedule'
   'Rather or very risky work'

B. Karasek, 1985. (The Job Content Questionnaire, 5-item version of the demand scale):
   'My job requires working very fast'
   'My job requires working very hard'
   'I am not asked to do an excessive amount of work'
   'I have enough time to get the job done'
   'I am free from conflicting demands that others make'

C. Johnson & Stewart, 1993; Johnson, Stewart, Hall, Fredlund & Theorell, 1996. (Job exposure matrix):
   'Is your job psychologically demanding?'
   'Is your job hectic?'

D. Bosma, Marmot, Hemingway, Nicholson, Brunner, & Stansfeld 1997. (The Whitehall II job demands scale):
   'Do you have to work very fast?'
   'Do you have to work very intensively?'
   'Do you have enough time to do everything?'
   'Do different groups at work demand things from you that you think are hard to combine?'

E. Lynch *et al*, 1997a, b.
The respondents were asked to rate how much mental strain or stress each of the following things caused them at work:
   'Excessive supervision of time schedules'
   'Troublesome supervisors'
   'Troublesome fellow workers'
   'Job responsibility'
   'Poorly defined tasks and responsibilities'
   'Risk of accidents'
   'Risk of unemployment'
   'Irregular work schedules'
   'Mental strain of work'
   'Work deadlines'
   'Physical strain of work'

F. van Yperen & Snijders, 2000.
   'Do you have to work very fast?'
   'Do you have too much work to do?'
   'Do you have to work extra hard to finish a task?'
   'Do you work under time pressure?'
   'Do you have to rush?'
   'Can you do your work in comfort?'
   'Do you have to deal with a backlog at work?'
   'Do you have too little work?'
   'Do you have problems with the pace of work?'
   'Do you have problems with the workload?'
   'Do you wish you could work at an easier pace?'

Figure 1.   Questions on psychological job demands from a number of scales used in studies on job demands and health, 1979–2002.

The point made by Steenland *et al*. (1997)—that job demands correlate positively with socio–economic status—is, however, a very important one. A large number of studies have found the same pattern, most notably the Whitehall II study (Marmot, Bosma, Hemingway, Brunner, & Stansfeld, 1997). Also, many recent studies on job strain and cardiovascular diseases (CVD) have been negative with regard to the expected association

between high demands and increased risk of CVD (Kristensen, 1999). The explanation for both of these empirical findings may be found by taking a closer look at the definition and measurement of the demand concept, which is the central point of this paper.

1.2. *Differential item functioning*

The concept of differential item functioning—which is defined and discussed in Section 1.3—will be the main methodological cornerstone of this paper. This makes it necessary to explain the methodological and theoretical implications of DIF for analyses of scales such as the scale for quantitative demands. It will be clear why DIF is important when we consider the reason for using scales instead of individual items (questions). Let us use a test in arithmetic as an example: if we want to test peoples' ability to solve arithmetical problems, we could choose to use only one item, such as $67 + 14 = ?$ A test based on only one item would, however, be quite crude, since it would result in a scale with only two values: correct and incorrect. The reliability would also be low since there would be a large element of chance (random error). Therefore we usually choose to have many items in a test since this will give more reliable and precise results. In principle we can ask about a large number of possible arithmetical problems, but for practical reasons we only use a very small fraction of these in a given test. The selection of items from the large item pool of those potentially possible is a process that is very poorly described by most researchers. This is a process where the intuition of the researcher and random circumstances play a very large—and often ignored—role. Now, suppose that women are better than men when adding and subtracting, while men are best when it comes to multiplying and dividing (a strictly hypothetical example!). This would have the consequence that a test involving many calculations with addition and subtraction would show that women are superior to men in the field of arithmetics, while a test involving many calculations with multiplication and division would show the opposite. That would be a clear case of DIF. Analyses of DIF have played a major role in psychometric testing in the field of intelligence and entrance tests since bias against racial groups, women, or disadvantaged socio-economic groups has been an important issue. For this reason DIF has also been called *item bias*. In connection with analyses of scales in the field of psychosocial factors at work, DIF has so far largely been ignored (a rare exception is Ørhede & Kreiner, 2000).

1.3. *A scale for demands at work without differential item functioning?*

Quantitative job demands are directly related to the *amount of work to be done*, and the basic source of stress is the possible mismatch between the amount of work and the time available to do it. One of the main research problems seems to be that the mismatch between time and task demands seems to manifest itself differently in different jobs. For example, an assembly line worker with fixed working hours may try to respond to increasing demands by working *faster*, while his boss will have little to win by talking faster on the phone. He will probably try to meet high demands by working *longer* hours. Thus, our main hypothesis is as follows:

Hypothesis: People in different jobs will have different possibilities for resolving (or trying to resolve) the mismatch between time and amount of work to be done, and this will lead to *differential item functioning* in scales on quantitative demands.

DIF is present in a scale if the response to an item for respondents at a given level on the scale differs for different groups of respondents (such as, for example, different age groups, men and women, or employees with different jobs); (Ørhede & Kreiner, 2000; Swaminathan & Rogers, 1990; Zumbo, 1999). DIF is a sign of low construct validity: a scale with a high degree of DIF may not measure what it is intended to measure in all parts of the population. If our hypothesis is confirmed, the conclusion will be that the results of the research depend on the type of questions that the investigator chooses to ask, not on the structures and processes that they want to study. It is quite clear that such a conclusion has far reaching consequences for the job demand literature.

In connection with the development of the Copenhagen Psychosocial Questionnaire (COPSOQ — see below), by two authors of the present paper (Kristensen and Borg), one of our aims was to operationalize the concept of quantitative demands in a 'pure' scale. By this we mean a scale without questions on role conflicts, emotional or other qualitative demands, and without 'psychological' items (such as 'my work is hectic'). However, in connection with our use of this new scale, we discovered that we had run into a new and unanticipated problem: DIF. Thus, our aims in this paper are: (1) to present a new scale for the measurement of quantitative demands at work and to compare this with other scales in the literature; (2) to analyse this scale with special attention to the problem of DIF; and (3) to propose solutions for measuring quantitative demands in future studies in the light of our results.

2. Method

2.1. Sample

Data for our analyses were obtained through a questionnaire study of a representative sample of all adult working Danes (age 20–60 years): The National Danish Psychosocial Work Environment Study. Names and addresses of a random sample were received from the national population registry. Respondents ($N = 1858$) were asked about a broad range of psychosocial work environment factors. The response rate was 62%; 49% of the respondents were women (Kristensen, Borg, & Hannertz, 2002).

2.2. Measures of job demands

One of the purposes of the original study was to develop a new questionnaire, the Copenhagen Psychosocial Questionnaire (COPSOQ; Kristensen, 2002). This questionnaire includes 30 scales on different aspects of the psychosocial work environment, health, stress, and well-being. Five of the scales measure different demands at work: quantitative demands (7 items, $\alpha = .80$); emotional demands (3 items, $\alpha = .87$); demands for hiding emotions (2 items, $\alpha = .59$); cognitive demands (8 items, $\alpha = .86$); and sensory demands (5 items, $\alpha = .70$). In the present paper we only analyse the scale on quantitative demands.

2.3. Measures of health and well-being

Health, fatigue, and psychological well-being were measured with three scales from the Short Form 36 (SF-36) questionnaire: General Health (5 items, $\alpha = .75$), Mental Health (5 items, $\alpha = .80$), and Vitality (4 items, $\alpha = .80$) (Ware, Snow, Kosinski, & Gandek, 1993) and three scales from the Setterlind Stress Profile: Behavioural Stress (8 items, $\alpha = .79$), Somatic Stress (7 items, $\alpha = .76$), and Cognitive Stress (4 items, $\alpha = .85$) (Setterlind &

Larson, 1995). Furthermore, the respondents were asked about their use of painkillers and sleeping pills during the last 3 months.

2.4. *Analyses*

DIF has often been analysed with regard to race, gender, socio–economic status, or age (Cammilli & Shepard, 1994; Holland & Wainer, 1993). In this paper we will concentrate on DIF with regard to the *job* of the respondent (in the following referred to as job-DIF). The 1858 respondents were categorized according to their main job at the time of the interview. We used the 5-digit coding system of the Danish version of the ISCO 1968 (International Standard Classification of Occupations). On the basis of this classification we grouped the respondents into 32 fairly homogeneous jobs such as driver, nurse, elementary school teacher, architect, etc. In this way we succeeded in classifying 1222 persons into 32 jobs with at least 20 respondents, while the remaining 636 had jobs with less than 20 respondents. In the job analyses below we only include the 1222 respondents with the 32 jobs.

All scales in COPSOQ have been given scale values of between 0 and 100. All items in the demand scales had five response options. (Wording of questions and response options are given in Table 1.) The items in the scales have been given equal weights and the response categories have been treated as equal interval categories. High levels on the scales mean high levels of the property being measured (such as, for example, high level of vitality, high stress level, or high quantitative demands). If more than half of the items in a scale were not answered, the person was considered missing on the scale. Hence, we used the conventions of the SF–36 questionnaire (Ware *et al.*, 1993).

We used factor analyses and analyses of internal reliability (Cronbach's α) according to normal standards and procedures in psychometric research (Nunnally & Bernstein, 1994; Streiner & Norman, 1998). Furthermore we used multivariate linear and logistic regression analyses and analyses of rank correlations (Spearman). According to the definition of DIF, response to a particular item should not depend on the person's job for respondents

Table 1. Basic characteristics of the 7–item scale on quantitative demands at work (Copenhagen Psychosocial Questionnaire).

Questions	Average score	*R* with total scale	Factor loading
a. Do you have to work very fast? ('work fast')	58.7	.44	.56
b. Is your work unevenly distributed so it piles up? ('work piles up')	49.8	.56	.65
c. How often do you not have time to complete all your work tasks? ('unfinished tasks')	34.2	.59	.72
d. Do you get behind with your work? ('behind')	30.0	.62	.73
e. How often can you take it easy and still do your work? ('take it easy')	52.7	.55	.71
f. Do you have enough time for your work tasks? ('enough time')	40.2	.60	.74
g. Do you have to do overtime? ('overtime')	44.8	.42	.47

$N = 1791$.
The response options: 'Always', 'Often', 'Sometimes', 'Seldom', 'Never/hardly ever' were given the scores of: 100, 75, 50 25 and 0 respectively. Reversed scores for items *e* and *f*.
Cronbach's α for the whole scale: .80.
Inter–item correlations: .22– .66.
Items *a* and *f* are from the Whitehall II Study (Bosma *et al.*, 1997), b is from the Finnish OSQ (Elo, Leppänen, Lindstrõin, Roppnen, 1992), *c* and *e* are from the Danish Agervold questionnaire (Agervold, 1998), *d* is from the Dutch QEEW questionnaire (van Veldhoven & Meijman, 1994), and *g* is from the QPSNordic (Dallner *et al.*, 2000).

experiencing the same level of quantitative demands (i.e. at the same level of the sum score). We tested this assumption using logistic regression of each item score on jobs where the sum score was entered as a covariate (Swaminathan & Rogers, 1990; Zumbo, 1999). The assumption of no DIF is then equivalent to independence between job and item score (i.e. a regression coefficient of zero) when the sum score is taken into account. For each item, we did an overall test of DIF across all jobs (entered as a class variable) and we evaluated the impact of DIF using explained variation (R^2) as suggested by Zumbo (1999). For the analyses reported here, we considered an item to exhibit severe DIF if the job explained more than 5% of the item variation, once the sum score was taken into account. We also examined DIF for each job, using the beta parameters from the logistic regression.

3. Results

We had originally included eight questions on quantitative demands in our test questionnaire when we developed the COPSOQ. One of the questions ('Do you have too little to do at work?') was excluded from the scale because of an extremely skewed response pattern, very low loading on the latent factor, and very low correlation with the other items. The remaining seven items seemed to function very satisfactorily as one scale on quantitative demands at work. The basic characteristics of this scale and the origin of the items are shown in Table 1. This scale lived up to our general 'rules of thumb' for scales in COPSOQ: theoretically meaningful, confirmed by factor analysis with all loadings above .40, Cronbach's α above .70, inter–item correlations between .20 and .70, all item correlations with total scale above .40, no strong ceiling or floor effects, and less than 5% missing values on all items. Furthermore, the average scores on the seven items were different, but still not too close to 0 or 100. The reliability of the whole scale ($\alpha = .80$) was clearly better than for most demand scales in the literature (Karasek *et al*., 1998; Kawakami, Kobayashi, Araki, Haratani, & Furui, 1995; Schwartz, Pieper, & Karasek, 1988).

A more critical look at the scale reveals a few minor problems already at this stage. Items *a* ('work fast') and *g* ('overtime') clearly have lower correlations (.42 and .44) with the whole scale than the other five items (.55 +) and the same two items have rather strong loadings on other demand scales. Item *a* ('work fast') has a loading of .40 on the scale of sensory demands, while item *g* ('overtime') has a loading of .37 on the cognitive demands scale. We were aware of these 'warning signs' when we constructed the scale but we chose to keep the two items in the scale. Our main reason was that we wanted to have a broad scale that would be able to capture many different signs of quantitative demands. We felt that important indicators of quantitative demand would be missing if these items were removed from the scale.

Table 2 shows the distribution of the 32 jobs on the scale of quantitative demands. The job averages go from 26 to 56 points on the 100–point scale. Most of the high–demand jobs appear to have been white collar jobs (managers, clerks, technicians, architects, etc.). Blue collar jobs known to have piece rate work or strict time limits such as food industry workers, cleaners, drivers, and cooks had low or rather low scores. Jobs in the human service sector also had low scores (nurses, kindergarten teachers, and home helps). Thus, the 7-item scale tended to paint a picture of high demands in a number of office jobs and low demands in the production sector, the service sector, and the human service sector. This picture is counter–intuitive and probably has low face validity for most researchers in the field of psychosocial work environment.

Our next step was to analyse job-DIF with regard to the seven items in the scale. All items showed significant DIF (all seven *p*-values were below .001). The most severe DIF

Table 2.   The mean score on the 7–item quantitative demands scale of 32 jobs in the National Danish Psychosocial Work Environment Study, (Range 0–100).

Rank	Job	Score
1	Managers	56.4
2	Systems planners	56.3
3	Secretaries	53.0
4	Bank clerk	52.2
5	High school teachers	51.3
6	Store managers	50.8
7	Warehouse assistants	49.7
8	Head clerks	48.9
9	Book keepers	48.5
10	Technicians	48.0
11	Architects	47.4
12	Office clerks	47.2
13	Electricians	46.3
14	Vocational school teachers	45.5
15	Nurses	45.0
16	Food industry workers	44.9
17	Elementary school teachers	44.6
18	Nurses' aids	43.7
19	Foremen	43.3
20	Cooks	43.1
21	Home helps	42.5
22	Mechanics	41.8
23	Drivers	40.9
24	Metal workers	40.6
25	Kindergarten teachers	39.9
26	Salesmen	39.8
27	Construction workers	39.3
28	Shop assistants	39.0
29	Agricultural workers	37.1
30	Cleaners	31.9
31	Foster parents	30.2
32	Kindergarten assistants	26.2

All groups: $N \geq 20$.

was seen for item a ('work fast') where job explained 6.3% of the variance and item g ('overtime') where job explained 8.3% of the variance. The rest of the items had explained variance between 2.4% and 4.9%. With regard to jobs we also found a fairly clear pattern: based on the logistic regression, three of the jobs showed strong DIF on four of the seven items—cleaners, food industry workers, and drivers. Figure 2 shows the pattern of DIF for these three jobs and also for three jobs with a low degree of DIF: mechanics, metal workers, and nurses.

In Figure 2 the response patterns for the six jobs are compared using the results of the logistic regression. A complete lack of DIF would result in bars that were close to zero. Figure 2 shows that cleaners answered very differently (taking the overall score into account) on the questions of the scale. The analyses showed job-DIF for five of the items for this group. On items b and g the scores were significantly lower than the overall job score, while they were higher on items a, e, and f. For the food industry workers there was significant job-DIF on four items, while for the drivers there was job-DIF on five items. Thus, for these three occupations the overall score on the scale on quantitative demands depended heavily on the selection of items. For mechanics, metal workers, and nurses we see no significant job-DIF.

Figure 2.   Illustration of differential item funtion (DIF) for a number of selected jobs: (a) three jobs with high job-DIF and (b) three jobs with low job-DIF.

The issue of job-DIF can also be illustrated by looking at the distribution of the 32 jobs on combinations of the seven items in the scale. This is illustrated in Figure 3, where we show two different combinations of items from the scale. The first part of the figure (3*a*) shows the distribution of the 32 jobs on items *c* ('unfinished tasks') and *d* ('behind'). The pattern is quite clear: jobs with high average scores on one of these items also have high average scores on the other (rank correlation = .84). All combinations of items *b*, *c*, and *d* gave similar pictures. This corresponds to a situation with low job-DIF. Figure 3(*b*) illustrates the opposite situation. In this case the two items are item *a* ('work fast') and item *b* ('work piles up'). Here the rank correlation between job averages was as low as .24.

We also wanted to elucidate possible differences with regard to the potential impact on health. In Table 3 we show the correlations (adjusted for age and gender) between the seven items and a number of health-related variables: the three scales from the SF-36 (General Health, Mental Health, and Vitality; Ware *et al*., 1993), the three scales from the Stress Profile (Cognitive Stress, Somatic Stress, and Behavioural Stress; Setterlind & Larson, 1995), and the use of two kinds of medication (pain-killers and sleeping pills). The seven items showed almost the same pattern with two exceptions: Item *g* ('overtime') was strongly associated with only one of the scales: behavioural stress. The correlations between item *a* ('work fast') and the scales were also comparatively low. The 7-item scale was strongly associated with all six scales measuring health and well-being. In contrast, all correlations with the use of medication were low (.00−.06).

Table 4 shows the same combination of variables, but here they are analysed in a multivariate linear regression analysis where gender and age were included as 'independent
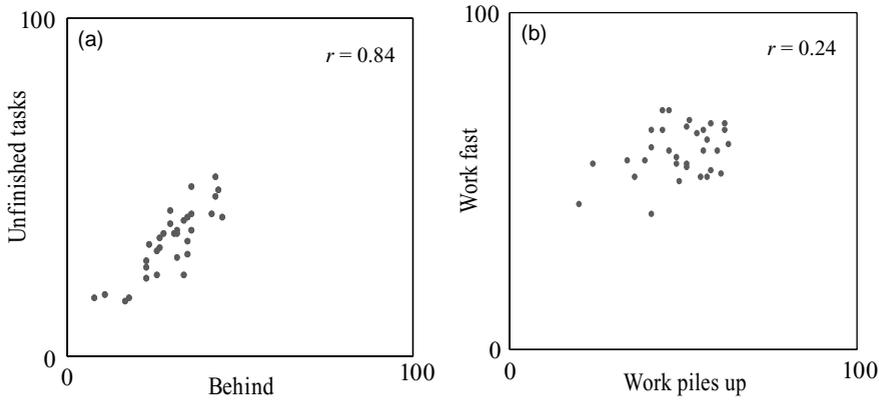
Figure 3.   The distribution of average scores of 32 jobs on two difference combinations of items on quantitative demands. (a) The distribution of average scores of 32 jobs on the two dimensions: 'Unfinished tasks' and 'Behind'; (b) The distribution of average scores of 32 jobs on the two dimensions: 'Work fast' and 'Work piles up'.

variables'. Item g ('overtime') was associated with Behavioural Stress in the expected direction, but an inverse association is seen with the scale for Vitality (high vitality is associated with high levels of overtime). For items a ('work fast'), c ('unfinished tasks'), and e ('take it easy') we only see one or no significant associations. Most of the significant associations in the table are seen for the three items b ('work piles up'), d ('behind'), and f ('enough time').

At this point in the analyses it was decided to construct a new and shorter scale with five items by omitting items a and g ('work fast' and 'overtime'). This decision was based on five different considerations: (1) *Theoretical* (see below). (2) *Reliability*. As indicated above, items a and g had the lowest correlations with the total scale, and the overall α was not reduced by removing the two items in spite of the fact that shorter scales usually have

Table 3.   Correlations between the seven items on quantitative demands, the 7-item scale, the 5-item scale of the COPSOQ and a number of health-related variables. Adjusted for age and gender.

| | Health-related variables | | | | | | | |
| | SF–36 scales | | | Stress scales | | | Use of medication | |
Items and scales	GH	MH	VT	Cognitive	Somatic	Behavioural	Pain-killers	Sleeping pills
a. 'work fast'	−.06*	−.13***	−.11***	.08**	.14***	.16***	.03	.04
b. 'work piles up'	−.08**	−.18***	−.16***	.17***	.14***	.24***	.05*	.05*
c. 'unfinished tasks'	−.05*	−.18***	−.16***	.18***	.11***	.19***	.00	.01
d. 'behind'	−.10***	−.23***	−.17***	.21***	.12***	.22***	.00	.04
e. 'take it easy'	−.10***	−.20***	−.18***	.14***	.15***	.19***	.03	.01
f. 'enough time'	−.13***	−.26***	−.28***	.24***	.21***	.26***	.06**	.03
g. 'overtime'	.00	−.08**	−.05	.09***	.07**	.17***	.01	.05*
7-item scale	−.10***	−.26***	−.23***	.23***	.19***	.30***	.03	.05*
5-item scale	−.12***	−.27***	−.25***	.25***	.19***	.30***	.03	.04

$N \geq 1791$.
*$p < .05$; **$p < .01$; ***$p < .001$.
Correlations are Spearman rank correlations.
Abbreviations of the SF–36 scales: GH: General Health. MH: Mental Health. VT: Vitality.

Table 4. Multivariate analyses of associations between the seven items of the quantitative demands scale of the COPSOQ and a number of health–related variables. Adjusted for age and gender.

| | Health–related variables | | | | | | | |
| | SF–36 scales | | | Stress scales | | | Use of medication | |
Items	GH	MH	VT	Cognitive	Somatic	Behavioural	Pain-killers	Sleeping pills
a. 'work fast'					.09★★★			
b. 'work piles up'			−.05★	.07★		.08★★		.29★
c. 'unfinished tasks'								
d. 'behind'		−.11★★★	−.06★	.12★★★		.07★★		
e. 'take it easy'		−.06★						
f. 'enough time'	−.13★★★	−.16★★★	−.27★★★	.14★★★	.17★★★	.16★★★	.12★★	
g. 'overtime'			.06★			.05★		

$N \geq 1791$.
★$p < .05$; ★★$p < .01$; ★★★$p < .001$.
All parameters in the table are standardized betas.

smaller αs. (3) *Factor analyses*. The same two items had the lowest loadings and had relatively high loadings on other scales. (4) *Job-DIF*. We found strong job-DIF for items *a* and *g*. (5) *Differential association pattern with potential endpoints*. We found that items *a*, *c*, *e*, and *g* showed weak or even 'opposite' associations with a number of health–related endpoints. While none of these points is sufficient alone, we felt that the overall evidence against keeping items *a* and *g* in our scale was very convincing.

The new and shorter 5–item scale had a Cronbach's α of .80 (the same as the longer scale, in spite of having fewer items), and inter-item correlations between .36 and .66. The correlations with the total scale varied from .53 to .64. In Table 5 we show the rank of the 32 jobs on the new 5–item scale and on the two omitted items. Comparisons between the three columns of the figure show some striking differences. For example, cooks rank as number 23 on the 5–item scale, as number 1 on item *a*, and as number 23 on item *g*. For food industry workers we find a very similar pattern. Architects rank as number 8 on the 5–item scale, as number 26 on item *a*, and as number 7 on item *g*. For other jobs we also find very different ranks: high school teachers (ranks 5 − 25 − 2), agricultural workers (29 − 22 − 7), home helps (17 − 12 − 31), head clerks (6 − 27 − 12), and drivers (28 − 8 − 5).

We then ran the job-DIF analyses on the new 5–item scale. In these analyses item *d* ('behind') did not show significant job-DIF when the number of tests was taken into account ($p = .011$, $R^2 = .017$) while the other items still showed significant job-DIF (all four test had $p < .005$, R^2 between .021 and .041). Fifteen jobs had no significant regression coefficients (against nine jobs for the longer scale). Thus, the level of DIF was substantially reduced but the new and shorter scale was not free from job-DIF.

The new scale showed strong associations with the six scales for health and psychological well-being (Table 3). Four of these correlations were 1−2 points higher with the 5–item scale than with the original 7–item scale, while two of the correlations remained unchanged. The correlations with use of medication remained practically unchanged. In a multivariate analysis with the new scale and the two single items as independent variables together with gender and age, the 5–item scale showed strong independent associations with the same six scales, while 'work fast' was independently associated with Somatic Stress, and 'overtime' with Vitality (inverse), Behavioural Stress, and the use of sleeping pills (Table 6).

Table 5. The distribution of 32 jobs on the 5-item scale on quantitative demands and on the two items: *a* ('work fast') and *g* ('overtime').

Points on the 5-item scale		Rank on the 5-item scale	Rank on item *a* 'work fast'	Rank on item *g* 'overtime'
53.5	Systems planners	1	14	6
50.8	Managers	2	5	1
50.7	Secretaries	3	9	22
47.9	Bank clerk	4	5	11
46.7	High school teachers	5	25	2
46.1	Head clerks	6	27	12
44.2	Book keepers	7	15	17
44.1	Architects	8	26	7
44.0	Office clerks	9	11	25
43.4	Technicians	10	13	9
43.3	Store managers	11	3	3
42.7	Warehouse assistants	12	4	4
41.7	Elementary school teachers	13	22	20
41.1	Electricians	14	15	9
40.8	Vocational school teachers	15	20	12
40.5	Foremen	16	22	24
40.0	Home helps	17	12	31
39.4	Nurses' aids	18	15	19
39.3	Nurses	19	5	16
37.6	Mechanics	20	28	15
37.1	Food industry workers	21	1	14
37.0	Kindergarten teachers	22	28	27
37.0	Cooks	23	1	23
35.9	Shop assistants	24	19	30
35.8	Metal workers	25	18	21
35.2	Salesmen	26	30	18
33.8	Construction workers	27	10	28
32.0	Drivers	28	8	5
30.8	Agricultural workers	29	22	7
28.4	Cleaners	30	20	32
25.9	Foster parents	31	32	25
21.1	Kindergarten assistants	32	31	29

$N \geq 20$ for all jobs.

4. Discussion

In this section of the paper we will first discuss a couple of general issues related to the concept and measurement of demands at work. Then we will discuss the results of our own analyses and continue with an attempt to develop a model for quantitative demands at work. On the basis of our theoretical and methodological considerations we will then make a number of recommendations regarding the measurement of demands in future studies.

4.1. *The issue of DIF: Different meanings or different conditions?*

As stated above, the present paper is, to our knowledge, one of the first systematic papers analysing job demand scales with regard to DIF (see also Ørhede & Kreiner, 2000). A number of authors have, however, been discussing the issue of DIF without referring to the concept. In the 1998 paper on the Job Content Questionnaire (JCQ), Karasek *et al.* wrote: 'The variability in the association of the psychological demands scale across samples supports the interpretation that its *meaning* may differ by population group' (Karasek *et al.*, 1998, p. 347), and Theorell *et al.* (1998) concluded along the same lines in another paper: 'The measures of psychological job demands need to be refined. Analyses in the case of

Table 6. Multivariate analyses of associations between the 5-item scale, two single items and a number of health-related variables. Adjusted for age and gender.

	Health-related variables							
	SF-36 scales			Stress scales			Use of medication	
Items and scale	GH	MH	VT	Cognitive	Somatic	Behavioural	Pain-killers	Sleeping pills
a. 'work fast'					.09★★★			
g. 'overtime'			.07★ (inverse)			.05★		.28★
5-item scale	− .12★★★	− .25★★★	− .28★★★	.25★★★	.15★★★	.24★★★		

$N \geq 1791$.
★$p < .05$; ★★$p < .01$; ★★★$p < .001$.
All parameters in the table are standardized betas.

psychological demands should perhaps be made separately in different social classes, since these questions may be perceived differently. Inferred psychological demands are particularly vulnerable to differences in social class' (Theorell *et al.*, 1998, p. 388).

While Karasek and Theorell write about differential *meaning* and *perception* by population groups and social classes, Hallqvist *et al.* (1998) go one step further:

In our group we have been concerned with the findings of lower prevalence of high demands among manual workers in cross-sectional surveys, and the main hypothesis has been that the measures are insensitive to demands experienced by manual workers. Questions such as 'Did you have enough time to complete your work tasks?' could for example be expected to be more apt for individuals who sell their competence rather than their time, an aspect of the theoretical foundation for separation of blue-collars and white-collars. The results here point in another direction. The measures of demands may if anything, be too unspecific with regard to non-manuals. And perhaps it is not only a measurement problem in that the concept of demands may be too loosely defined to capture only what is really demanding in this social stratum (Hallqvist *et al.*, 1998, p. 1414).

We reprint this rather long quote from the paper by Hallqvist *et al.* because the line of thought presented here is very close to the main points of this paper: DIF in scales on demands at work is not only a question of perception and meaning but reflects the basic structural and material conditions of different classes and occupational groups. The comments by Hallqvist *et al.* (1998) have been an important source of inspiration for this paper.

A temporary conclusion is that the job demand concept seems to be both poorly defined and measured. A number of the main researchers in the field seem to agree with us on this issue: 'Clearly, one practical conclusion of this study is that the basic measurement of the demand construct should be improved' (Johnson *et al.*, 1996, p. 329); 'Exposure measurement should be improved, especially for the psychological demand variable' (Steenland *et al.*, 1997, p. 260), and 'The measures of psychological job demands need to be refined' (Theorell *et al.*, 1998, p. 388).

4.2. Discussion and integration of the results of the present paper

Two of the main advantages of our study are that the study base is a large heterogeneous sample, and that so many different demand dimensions were included in the questionnaire. We would, however, also like to mention two of the weaknesses of the study before we proceed to the results. First, the job classification was rather crude. We wanted to consider as many jobs as possible, and we also wanted the number of respondents for each job to be

at least 20 in order to limit the random variation. The consequence was that we formed a number of job specifications that were rather broad. For instance, 'drivers' included bus drivers, truck drivers, and taxi drivers. A larger study base would have made it possible to distinguish between such subgroups. We believe that the crude job classification tends to blur the degree of job-DIF. The other problem is that we only had cross-sectional data. This means that our analyses in Tables 3, 4 and 6 should not be interpreted as analyses of causal associations. We bring these tables in order to show the differential picture, not to demonstrate causality.

The main conclusion of our analyses seems to be the following: The (7-item) scale for quantitative demands at work was confirmed by analyses of internal reliability and by factor analyses, but the scale is still unsatisfactory due to job-DIF. Our analyses of job-DIF have clearly demonstrated that the rank of a job on a scale depends strongly on the selection of items for the scale. If we had constructed a scale with many questions of the 'overtime' type, jobs such as high school teachers, drivers, store managers, and architects would have received high job demand scores. Had we, on the other hand, constructed a scale with many questions on 'fast work pace', the result would have been quite different. In this case cooks, food industry workers, and nurses would have received high scores, while high school teachers and architects would have been close to the bottom. Clearly, these items cannot measure the same dimension in spite of the fact that we find satisfactorily high correlations between them at the individual level.

It might be argued that job-DIF would be a minor problem if all the items in the scale measured job factors with the same health effects. In that case it would be like adding different types of fruits in a questionnaire on dietary habits. The problem is that we have no evidence for this. Our cross-sectional analyses in Tables 3 and 4, and 6 seem to indicate that the health effects are different, and our knowledge of the literature on fast-paced work versus long working hours seems to suggest the same. For instance, it does make sense that 'work fast' is related to somatic symptoms, 'work piles up' to the use of sleeping pills, and 'behind' and 'enough time' to (poor) mental health and cognitive stress (Table 4). Prospective studies are needed to elucidate this issue further.

When we decided to split the concept of job demands into five different dimensions in the COPSOQ we thought that we had solved the problem of heterogeneity of the demand construct. We can now see that we need to refine our measures further. The two main dimensions of quantitative demands at work seem to be *intensity* (work pace) and *extensity* (number of working hours). We will proceed with this issue in the next paragraph.

4.3. *How should quantitative demands be measured?*

We think that having two scales for quantitative demands at work could be a possible solution to the problems discussed in this paper. The first could be a basic scale for quantitative demands such as the 5-item scale presented in this paper. This scale measures different aspects of the mismatch between task and time, and it is our impression that it is valid and has good psychometric qualities. Needless to say, it has to be tested in more studies, in particular prospective studies. With regard to intensive quantitative demands we think that our item *a* ('work fast') should be supplemented with other questions on work pace and tempo. This is a task for future development of the COPSOQ or other questionnaires. With regard to the other basic dimension—extensity—our solution is different. We believe that the issue of working hours should be elucidated by asking a number of specific questions and not by forming a scale.
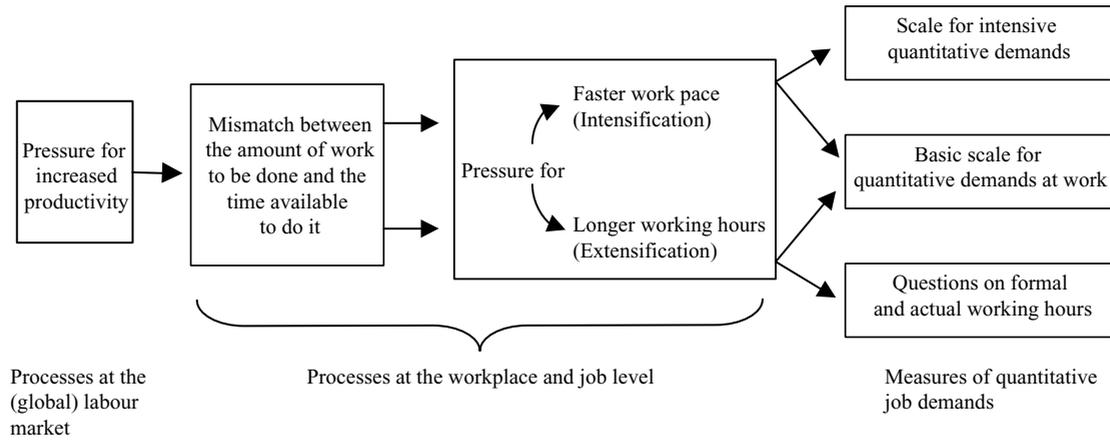
Figure 4.   Model for the measurement of quantitative job demands.

With regard to the problems of measuring demands, the 'three-factor solution' that we are suggesting is illustrated in Figure 4: the pressure for increased productivity leads to a mismatch between the amount of work and the time available to do it. This mismatch leads to pressure for both longer working hours and faster work pace. These two ways of increasing productivity will be applied differently in different segments of the labour market depending on laws, agreements, wage structure, professional norms, etc. Consequently, quantitative demands could be measured in three ways: a scale for *intensive* demands, a scale for the *basic mismatch between task and time*, and *specific questions on working hours*. We are not claiming that this solution solves all the problems with job-DIF, but we do think that our suggestion represents an improvement relative to the scales that are being used today (such as those shown in Figure 1). Most of these scales include a mixture of intensity, extensity, and items that measure mismatch between task and time. For instance, the JCQ (scale B in Figure 1) includes an item on intensity ('working very fast'), an item on extensity ('excessive amount of work'), and an item on mismatch ('enough time to get the job done'). Furthermore, this scale also has an item ('working very hard'), which seems to refer to physical as well as psychological strain, and an item on role conflict ('conflicting demands'). The other scales shown in Figure 1 also appear to contain internal heterogeneity, which is a strong indicator of potential problems with DIF in these scales.

We also recommend that analyses of DIF in the important scales of a study should always be included in research relying on questionnaire scales. If DIF is present—with regard to age, gender, class or occupation—knowledge about it will always give more insight than overlooking this important information.

5. Conclusions

In this paper we have been discussing one of the main dimensions of modern research on psychosocial working conditions: quantitative job demands. This dimension has been in focus for at least 50 years of research and it is our belief that job conditions in the global economy will be characterized by increasing quantitative job demands for many years to come. Hence, the issue of quantitative job demands is important and relevant.

One of our main conclusions is that (quantitative) job demands have been poorly conceptualized, defined and measured in the literature so far. In most cases researchers have used demand scales with poor quality and unknown validity. Most researchers have been satisfied with using the same scales as others have used. Very few have taken a close look at the very items of the scales employed.

A number of studies on job strain and cardiovascular disease and other studies on job demands and health, have shown negative or contradictory results during the last 10–15 years. We believe that one of the main explanations may be the use of job demand measures with low validity. These measures have to a large degree measured white collar demands associated with job conditions of higher socio-economic classes. This type of measurement bias has been mentioned by a few authors but the proper ways of analysing the issue and suggestions for solving the problem have been lacking. In this paper we have used analyses of differential item functioning in order to elucidate the issue and we have suggested the use of refined scales in the hope of improving future research in this important field.

References

AGERVOLD, M. (1998). *Spørgeskema til psykosocialt arbejdsmiljø—kortlægning og ændring*. [Questionnaire for psychosocial work environment—survey and change.] Copenhagen: Arbejdsmiljøfondet.

ALFREDSON, L., KARASEK, R., & THEORELL, T. (1982). Myocardial infarction risk and psychosocial work environment: An analysis of the male Swedish working force. *Social Science and Medicine*, *16*, 463–467.

BOSMA, H., MARMOT, M. G., HEMINGWAY, H., NICHOLSON, A. C., BRUNNER, E., & STANSFELD, S. A. (1997). Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study. *British Medical Journal*, *314*, 558–565.

CAMMILLI, G. & SHEPARD, L. A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks: Sage.

DALLNER, M., ELO, A.-L., GAMBERALE, F., HOTTINEN, V., KNARDAHL, S., LINDSTRÖM, K., SKOGSTAD, A., & ØRHEDE, E. (2000). *Validation of the General Nordic Questionnaire (QPSNordic) for Psychological and Social Factors at Work. Nord 2000:12*. Copenhagen: Nordic Council of Ministers.

DE JONGE, J. & KOMPIER, M. A. J. (1997). A critical examination of the demand-control-support model from a work psychological perspective. *International Journal of Stress Management*, *4*, 235–258.

ELO, A.-L., LEPPÄNEN, A., LINDSTRÖM, K., & ROPPONEN, T. (1992). *OSQ. Occupational Stress Questionnaire: User's Instructions*. Helsinki: Finnish Institute of Occupational Health.

HALLQVIST, J., DIDERICHSEN, F., THEORELL, T., REUTERWALL, C., AHLBOM, A., & SHEEP STUDY GROUP. (1998). Is the effect of job strain on myocardial infarction risk due to interaction between high psychological demands and low decision latitude? Results from Stockholm Heart Epidemiology Program (SHEEP). *Social Science and Medicine*, *46*, 1405–1415.

HOLLAND, P. W. & WAINER, H. (Eds.). (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.

JOHNSON, J. V. & STEWART, W. F. (1993). Measuring work organization exposure over the life course with a job-exposure matrix. *Scandinavian Journal of Work, Environment & Health*, *19*, 21–28.

JOHNSON, J. V., STEWART, W., HALL, E. M., FREDLUND, P., & THEORELL, T. (1996). Long-term psychosocial work environment and cardiovascular mortality among Swedish men. *American Journal of Public Health*, *86*, 324–331.

KARASEK, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, *24*, 285–308.

KARASEK, R. A. (1985). *Job Content Questionnaire and User's Guide*. Lowell: Department of Work Environment, University of Massachusetts Lowell.

KARASEK, R. A., BRISSON, C., KAWAKAMI, N., HOUTMAN, I., BONGERS, P., & AMICK, B. (1998). The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology*, *3*, 322–355.

KARASEK, R. & THEORELL, T. (1990). *Healthy Work. Stress, Productivity, and the Reconstruction of Working Life*. New York: Basic Books.

KAWAKAMI, N., KOBAYASHI, F., ARAKI, S., HARATANI, T., & FURUI, H. (1995). Assessment of job stress dimensions based on the job demands-control model of employees of telecommunication and electric power companies in Japan: Reliability and validity of the Japanese version of the Job Content Questionnaire. *International Journal of Behavioral Medicine*, *2*, 358–375.

KRISTENSEN, T. S. (1999). Challenges for research and prevention in relation to work and cardiovascular disease. *Scandinavian Journal of Work, Environment & Health*, *25*, 550–557.

KRISTENSEN, T. S. (2002). A new tool for assessing psychosocial factors at work: The Copenhagen Psychosocial Questionnaire. *TUTB Newsletter* [Newsletter of the European Trade Union Technical Bureau for Safety and Health], *19–20*(September), 45–47.

KRISTENSEN, T. S., BORG, V., & HANNERTZ, H. (2002). Socioeconomic status and psychosocial work environment. Results from a national Danish study. *Scandinavian Journal of Public Health*, *30*, 41–48.

LYNCH, J., KRAUSE, N., KAPLAN, G. A., SALONEN, R., & SALONEN, J. T. (1997a). Workplace demands, economic rewards, and progression of carotid atherosclerosis. *Circulation*, *96*, 302–307.

Lynch, J., Krause, N., Kaplan, G. A., Tuomilehto, J., & Salonen, J. T. (1997b). Workplace conditions, socioeconomic status, and the risk of mortality and acute myocardial infarction: The Kuopio Ischemic Heart Disease Risk Factor Study. *American Journal of Public Health*, *87*, 617–622.

Marmot, M. G., Bosma, H., Hemingway, H., Brunner, E., & Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet*, *350*, 235–239.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw–Hill.

Ørhede, E. & Kreiner, S. (2000). Item bias in indices measuring psychosocial work environment and health. *Scandinavian Journal of Work, Environment and Health*, *26*, 263–272.

Schwartz, J. E., Pieper, C. F., & Karasek, R. A. (1988). A procedure for linking psychosocial job characteristics data to health surveys. *American Journal of Public Health*, *78*, 904–909.

Setterlind, S. & Larson, G. (1995). The Stress Profile: A psychosocial approach to measuring stress. *Stress Medicine*, *11*, 85–92.

Steenland, K., Johnson, J., & Nowlin, S. (1997). A follow-up study of job strain and heart disease among males in the NHANES1 population. *American Journal of Industrial Medicine*, *31*, 256–260.

Streiner, D. L. & Norman, G. R. (1998). *Health Measurement Scales. A Practical Guide to their Development and Use*, 3rd edn. Oxford: Oxford University Press.

Swaminathan, H. & Rogers, J. H. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*, 361–370.

Theorell, T., Tsutsumi, A., Hallqvist, J., Reuterwall, C., Hogstedt, C., Fredlund, P., Emlund, N., Johnson, J. V., & the SHEEP Study Group (1998). Decision latitude, job strain, and myocardial infarction: A study of working men in Stockholm. *American Journal of Public Health*, *88*, 382–388.

van der Doef, M. & Maes, S. (1999). The job demand-control (-support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress*, *13*, 87–114.

van Veldhoven, M. & Meijman, T. F. (1994). *Questionnaire on the Experience and Assessment of Work*. Amsterdam: The Foundation for Quality in Occupational Health Care.

van Yperen, N. W. & Snijders, T. A. B. (2000). A multilevel analysis of the demand-control model: Is stress at work determined by factors at the group level or the individual level? *Journal of Occupational Health Psychology*, *5*, 182–190.

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey. Manual and Interpretation Guide*. Boston: New England Medical Center, The Health Institute.

Zumbo, B. D. (1999). *Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Ottawa, ON: Department of National Defense, Directorate of Human Resources and Evaluation.