Measuring psychosocial job strain with the job content questionnaire using experienced job evaluators

Ostry, AS, MA, M.Sc, Ph.D.; Marion, SA, MD, M.Sc; Demers PA. Ph.D.; Hershler R, M.Sc; Kelly S, B.Sc; Teschke K, PhD; Hertzman C, MD, MSc.

All the authors from Department of Health Care and Epidemiology, UBC

Address correspondence to:

Aleck Ostry
Dept. of Health Care and Epidemiology
5804 Fairview Avenue,
Vancouver BC
V6T 1Z3

Phone: 604-822-5872
Fax: 604-822-4994

e-mail: ostry@unixg.ubc.ca

Short Title: Measuring psychosocial job strain with experts
ABSTRACT

Background

A union/management system of job evaluation has been in place in the British Columbia (BC) sawmill industry since the late 1960s. This system uses an instrument, very similar to the Job Content Questionnaire (JCQ) to evaluate psychosocial work conditions for sawmill jobs.

Methods

Four experienced evaluators, one from the union and three from industry, independently estimated psychosocial work conditions for 54 current job titles in a “typical” coastal sawmill using a shortened, 18-question version of the JCQ questionnaire.

Results

Inter-rater reliability was acceptable for control but not for co-worker social support, physical demand, or psychological demand. Reliability was least for psychological demand.

Conclusions

Experienced job evaluators in the sawmill industry were able to reliably estimate only the control dimension of the JCQ. The observed lowest reliability for psychological job demand may be due to the imprecise construct definition in the domain of the JCQ instrument.

KEY WORDS: Psychological demand, control, expert raters, sawmills.
INTRODUCTION

Over the past 20 years a number of researchers have raised concerns regarding the reliance on self-reports of psychosocial work conditions for measuring job strain (Frese and Zapf, 1988; Kasl, 1978, 1987, 1998). In particular, affective reactions to a job might influence perception and reporting as might personality traits such as negative affectivity (Chen et al., 1995; Spector, 1989). As well, self-reports are a particular problem if the outcome is also measured with a self-report, because of what Kristensen calls the triviality trap or common methods variance (Kristensen, 1995, 1996). In this case, there may be contamination between measures.

In a number of American (Alterman et al., 1994; Karasek et al., 1988; Pieper et al., 1989; Reed et al., 1989) and Swedish studies (Alfredson et al., 1985; Johnson et al., 1989, 1993, 1996; Karasek et al., 1981) self-reports were pooled across occupations in order to increase the objectivity of measures of job characteristics. As well, expert raters have been used in at least six investigations (Baneryd, 1974; Frankenhaeuser and Gardell, 1976; Johannson et al., 1978; North et al., 1996; Stansfeld et al., 1993; Bosma et al., 1997). Two of these, using a single expert, were of sawmill workers in Sweden (Baneryd, 1974; Frankenhaeuser and Gardell, 1976; Johannson et al., 1978).

Most of these investigations were based on direct expert observation and estimation of the work tasks. However, in the Whitehall study, a different expert method was used as groups of experts reviewed the job descriptions of workers directly under their supervision and estimated psychosocial work conditions using the JCQ. One hundred and
forty supervisors in 18 different departments used the JCQ to estimate psychosocial work conditions for the jobs of 8,838 civil servants. In this investigation, “men and women with self-reported low job control had about 1.5 to 1.9 times the risk of developing new CHD compared with participants with high control. Similar results were found with externally assessed job control” (my italics) (Bosma et al., 1998).

Because expert ratings have not been used extensively to estimate psychosocial work conditions, it is not clear who the best possible expert raters might be or the best way to conduct such ratings. However, many industries employ experts whose work is job evaluation. The purpose of job evaluations vary but generally they are used to estimate job function, demand, pacing, effort, concentration, responsibility, exposure to hazards, and related conditions of work in order to establish pay rates for groups of jobs within an industry.

Where they exist, there are some advantages of using job evaluators as expert raters for psychosocial job conditions. First, they are likely to know, from years of direct observation, enough about work conditions to effectively estimate demand, control, and social support across an entire industry or segment of an industry. Second, unlike personnel managers and supervisors they are not in the line of authority, so are more likely to rate the job than rate the person or persons they know who fill these jobs. Third, because some industries have joint union/management teams conduct these job evaluations, there may be, by using pooled assessments from both sides, a chance to reduce management or union bias in scores.
The purpose of this investigation was to develop a method for expert rating of psychosocial job conditions using a panel of expert job evaluators and to test the reliability of their assessments using a modified version of the JCQ.

**MATERIALS AND METHODS**

A union/management system of job evaluation has been in place in the BC sawmill industry since the late 1960s. This job evaluation system relies on an instrument developed within the industry to measure psychological and physical demand, control over skill use, control over decision making, and physical and other hazards to which workers may be exposed. This instrument has conceptual congruence with the JCQ (Karasek, 1985). The expert evaluators are therefore familiar with dimensions used in the JCQ and also with applying these to assessments of sawmill job titles.

Since the 1960s, a total of six job evaluators (with over 20 years’ experience in sawmill job evaluation in BC) have been employed in the sawmill job evaluation program and were potential interviewees. Three were currently employed by industry, one was currently employed by the union and two were recently retired from the union. All three industry raters agreed to participate as did the currently employed union expert. The 4 job evaluators were a convenience sample of job evaluators from the BC sawmill industry. One of the retired union experts was too ill to participate and the other refused without giving a reason. Each of the 4 participants had over 30 years’ experience. They were blinded to the purpose of the study but were told that it was part of a broad investigation of the impact of new technology in the industry.
Using wage rate charts, a list of 54 of the most common job titles in the industry was developed by the research team. Each rater was asked to independently estimate exposure for current job titles in a “typical” coastal sawmill, using a shortened, 18-question version of Karasek's questionnaire (Appendix A). Although the questionnaire was shortened in order to reduce interviews to a manageable length of time each interview still took approximately 2.5 hours. To test intra-rater reliability, 10 percent of jobs were randomly selected from the job title list and re-presented to the raters towards the end of the interview so that 54 core and 5 repeat job titles were rated for a total of 59 job titles.

For the 18-item version of the JCQ used in this investigation, four dimensions were measured: control (nine items), psychological demand (five items), physical demand (one item), and co-worker social support (two items). With the exception of co-worker social support all questions were derived from Karasek's original JCQ (Karasek, 1985). The two questions on co-worker social support were obtained from the Swedish version of Karasek's questionnaire (Johnson et al., 1989). Finally, a question on noise exposure was added, because of its importance as a sawmill exposure. This question was developed in consultation with occupational epidemiologists in the sawmill cohort study group at the University of British Columbia.

The purpose of this study was to find out whether or not the raters’ mean estimate of exposure was reliable. That is, had we selected different groups of raters, would the group means of their estimates have been reliable? The measure that best estimates this
form of reliability is the intraclass correlation coefficient for all raters (ICC 2,4) termed the “group” ICC in this paper and designated ICC (2,k) by Shrout and Fleiss (1979).

As well, the intraclass correlation coefficient for individual raters (ICC 2,1) termed the “individual” ICC in this paper (Shrout and Fleiss, 1979) is used to measure intra-rater reliability. Finally, for each of the 5 items in the psychological demand construct, the individual ICCs (termed ICC (2,1) by Shrout and Fleiss) were estimated for exposures based on that item only. The purpose was to assess inter-rate reliability of each item individually. The component of variance distributed between raters, job titles, and the residual is also presented.

**RESULTS**

Intra-rater reliability is shown in Table I. According to Nunnally (1978), an ICC of 0.90 is the minimum standard for reliability. Results indicate that ICC’s were below this standard and that intra-rater agreement for all items was poor.

Table II describes inter-rater reliability in terms of the proportion of the variance explained between job titles, between raters, and residual. In all cases, the “between job title” variance exceeded the “between raters” variance, but by varying degrees. These factors are summarized by the Group ICCs which were between 0.63 and 0.92. Agreement was acceptable for control, but less than Nunnalley’s 0.90 standard for other dimensions of the JCQ.
Inter-rater reliability was worst for psychological demand. In order to explore this further, the reliability for each of the five items in the psychological demand construct was determined (Table III). The items “not excess” and “enough time” were much less reliably estimated than the remaining 3 items.

**CONCLUSIONS**

The purpose of this study was to determine whether a group of experts could reliably estimate psychosocial work conditions. In fact, the four expert job evaluators were able to estimate control reliably. The dimensions of the JCQ; co-worker social support, physical demand, noise, and psychological demand were not estimated reliably.

In most studies with the JCQ control is usually the most reliably determined variable and psychological demand the least. Using self-reports for three American quality of employment surveys for 2,946 men (in 1969, 1972, and 1977) Karasek found the between-occupation variance for psychological demand was 4.2 percent versus 34.7 percent for control and 25.9 percent for physical demands (Karasek, 1990; Schwartz et al., 1988).

Why was the reliability for psychological demand low relative to the other dimensions measured in this investigation? In a study estimating psychosocial work characteristics using several panels of experts in a single workplace inter-rater agreement was higher for moderators compared to stressors (Frese and Zapf, 1988). According to these authors “the resources (moderators like control) can be well observed and without much interference from cognitive and emotional processing; in contrast, the judgement of
stressors is more private and therefore more strongly related to cognitive and emotional processing. Therefore, the average inter-rater agreements are lower for the stressors” (Frese and Zapf, 1988, 382). This observation implies that psychological demand, because of its larger cognitive/emotional component relative to moderators, will always be less reliably estimated by external observers because it is more opaque to external scrutiny than other dimensions of job strain.

Expert assessment of psychological demand may also be hampered due to the relative lack of construct validity for this dimension. For example, it has been noted that the two psychological demand items “work hard” and “work fast”, particularly when estimated for blue-collar industrial jobs, likely measure physical as well as psychological demand (Karasek et al., 1998; Landsbergis et al., 2000). And, in a more general sense, as noted by Karasek in a discussion of cross-national studies using the JCQ, “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).

While it may be intrinsically difficult for expert observers to estimate psychological demand, because of its lack of construct validity, it is difficult to know which of these explanations best accounts for the low reliability scores for psychological demand. Clearly, the psychological demand construct in the JCQ requires improvement.

The newest version of the JCQ contains four new psychological demand questions which may improve the construct validity of this dimension. However, this present investigation
indicates that the two questions measuring whether work is “excessive” and whether a worker has “enough time” to complete a job (which are retained in the new version) were the least reliably measured items. According to the interviewer employed to carry out this study, the negative wording of the questions “the job does NOT involve excessive work” posed difficulties for the raters. The reliability of this item might be increased if worded positively.

It is not clear why the question of whether a worker has enough time to do a job was difficult to measure reliably while the question on conflicting demand was measured very reliably. Expert raters estimating psychosocial job conditions in a “generic” versus an actual sawmill will most reliably answer questions that are focused on job structure rather than the worker. It may be that the structure of a specific job has consistent attributes on the dimension of conflicting demands and that no matter what sawmill the job is located in, that in general, this dimension can be assessed by outside experts. However, whether a worker has enough time to do a job may be more a measure of the person or workplace specific factors such as line speed, deadlines, and staffing levels which were not assessed by expert job evaluators in this study.

There are several major limitations to this investigation. First, it may not be possible to blind evaluators (although it is unclear that this is necessary or desirable) and second, it may not be possible to find job evaluators in other industries. This point is germane when considering the generalizability of the method. Third, while these objective estimations of psychosocial job conditions using a generic sawmill (that is without direct reference to
or observation of specific plant-level conditions) have produced reliable estimates, this may be partly because of artificial consensus among raters working together in the same job evaluation program.

In summary, this investigation contributes to the oft stated need for the development of expert methods for measuring job strain. The study shows that expert job evaluators were able to assess three dimensions of the JCQ (control, co-worker social support, and physical demands) with moderate to substantial reliability. Reliability of expert assessments for psychological demand were fair. This could be due to problems of construct validity rather than failures in the expert method tested in this investigation.

REFERENCES


APPENDIX A (Modified JCQ)

b. * The job involves a lot of repetitive work
c. * The job requires a high level of skill
d. * The job has a variety of tasks
e. * The job requires creativity
f. * The job allows opportunity for the worker to develop his own special abilities
g. * The worker has a lot to say about what happens on the job
h. * On this job, the worker has a lot of freedom to decide how to do the work
i. * The job allows the worker to make a lot of decisions on his own
j. ** The job requires working very fast
k. ** The job requires working very hard
l. ** The job does not involve an excessive amount of work
m. ** The worker has enough time to get the job done
n. ** The job is free from conflicting demands
o. *** The job requires lots of physical effort
p. **** The worker can leave this job to talk with co-workers.
q. **** The worker can interact with co-workers while they work
r. The job is noisy

* = Control items; **=Psychological demand items; ***= Physical demand items; ****=Social support items
**Table I: Intra-rater reliability of 4 sawmill job evaluators expressed as individual ICCs**

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Psy.</th>
<th>Physical</th>
<th>Social</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=5</td>
<td>Rater #1</td>
<td>0.66</td>
<td>0.07</td>
<td>1.0</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Rater#2</td>
<td>0.69</td>
<td>0.68</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Rater #3</td>
<td>0.70</td>
<td>-0.50</td>
<td>0.43</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Rater#4</td>
<td>0.96</td>
<td>0.66</td>
<td>1.0</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.75</td>
<td><strong>0.23</strong></td>
<td><strong>0.73</strong></td>
<td><strong>0.54</strong></td>
<td><strong>0.54</strong></td>
</tr>
</tbody>
</table>

Psy = Psychological demand.
Table II: Inter-rater reliability for 4 sawmill job evaluators

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>Control</th>
<th>Psy.</th>
<th>Physical</th>
<th>Social</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Job Titles</td>
<td>80.0</td>
<td>42.0</td>
<td>63.0</td>
<td>70.0</td>
<td>54.0</td>
</tr>
<tr>
<td>Between Raters</td>
<td>1.0</td>
<td>11.0</td>
<td>3.0</td>
<td>2.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Residual</td>
<td>19.0</td>
<td>47.0</td>
<td>34.0</td>
<td>28.0</td>
<td>39.0</td>
</tr>
<tr>
<td>INDIVIDUAL ICC</td>
<td>0.73</td>
<td>0.26</td>
<td>0.52</td>
<td>0.60</td>
<td>0.41</td>
</tr>
<tr>
<td>GROUP ICC</td>
<td>0.92</td>
<td>0.63</td>
<td>0.82</td>
<td>0.86</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Psy. = Psychological demand.
Table III: Inter-rater reliability among 4 sawmill job evaluators for components of psychological demand

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>Fast</th>
<th>Hard</th>
<th>NotExcess</th>
<th>Enough</th>
<th>Conflict</th>
<th>Time</th>
<th>Demands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Job Titles</td>
<td>62</td>
<td>48</td>
<td>29</td>
<td>26</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Raters</td>
<td>3</td>
<td>0</td>
<td>23</td>
<td>6</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>35</td>
<td>52</td>
<td>48</td>
<td>68</td>
<td>35.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDIVIDUAL ICC</td>
<td>.50</td>
<td>.31</td>
<td>.13</td>
<td>.04</td>
<td>.53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fast= The job requires working very fast; Hard=The job requires working very hard; Not excess=The job does not involve an excessive amount of work; Enuf time=The worker has enough time to get the job done; Conflict demand=The job is free from conflicting demands.