APPLICATION OF INDUSTRIAL-STRENGTH WORKLOAD ASSESSMENT TOOLS

Ged Morrisroe

Ergonomics Group, Network Rail, 40 Melton Street, London, NW1 2EE

Keywords: mental workload, tools, signaller, railway, assessment, ergonomics.

Abstract

This paper addresses the experiences gained through administering the mental workload toolset, developed for railway use by the Institute for Occupational Ergonomics at University of Nottingham, on behalf of Network Rail.

The tools have been applied to the signalling tasks for locations which have been referred to the assessment team. The assessment team has comprised operational experts and an ergonomist.

The efficacy of these tools and representativeness of the results is discussed.

1 Background

The subject of occupational workload has a long and distinguished reputation as a theoretical and practical tar pit. Researchers that stray into this field emerge many years later with either another grand unified theory on the subject (subtly different to the ones already in the expanding list) or with significant reservations and qualification of the subject area. In fact, the subject seems to be stuck and progress over the years seems to be ephemeral.

In industry, workload is manifest and recognisable. Logically its effects on the individual and on the performance of the overall systems at work cannot be ignored even if there is lack of a single, sharp, pithy and universally agreed definition and theoretical underpinning for the subject.

So, this paper sidesteps the embroiled theory and concentrates on reporting the experience thus far of applying a set of workload assessment tools developed for the rail industry and reported elsewhere in this book in chapters by Wilson et al, Pickup et al and Mitchell et al.

This paper lists the toolset elements that have been applied to date, and for each reports the experience of applying them within signalling centres operated by Network Rail. Any refinements that we have made are described and a resume of the observations about the tools is presented.
The title of this paper refers to ‘industrial strength tools’ for workload assessment (at higher levels of workload - underload has not been addressed in these particular assessments). The requirements for such industrial strength or industrial suitability can be stated thus:

- detect workload and indicate whether it is excessive,
- identify the sources of demand and the drivers that are associated with the perception of excessive effort,
- can be administered within a half day site visit.

The success of the tools in meeting these requirements will be evaluated, using the experiences and outcomes of real use made by a team that does not include any of the original tool set developers.

2 The Network Rail Workload Assessment Toolset

The context in which these tools have been applied has been to address the question of whether or not a workload problem exists for a signaller. If it is deemed to be the case, then can the tools determine the sources and factors associated with the workload?

Process

To date, the source of requests for a workload assessment have arisen as a result of staff complaint, signalling managers’ requests or by referral from a third party (eg an Investigation). In each case a small team from the Ergonomics Group at Network Rail, comprising at least one operational specialist and an ergonomist, have carried out the assessment.

The team arrange a mutually convenient time to undertake the assessment and plan to be present during a significant time period that coincides with that of the perceived workload. On the day of the assessment, after introducing ourselves and the methods, we administer all or part of the workload toolset. We make it clear to the signallers participating that the assessment can be terminated at any time and for any reason at their request. Likewise, if the operations specialist deems that the assessment should be terminated for operational reasons then we leave. This report is based on about 10 assessments, none have been terminated prematurely, and only a small number have required a follow-up visit to check on the reliability of the findings by reviewing them with a different signaller.

The Toolset

The following tools have been used in real workload assessments:

1. Operational Demand Evaluation Checklist (ODEC)
2. Integrated Workload Scale (IWS) and Timeline Analysis
3. Workload Principles
4. SWAT (Subjective Workload Assessment Technique)
5. Rail Adapted Modified Cooper Harper (RMCH)
6. Workload Probes (decision ladders)

Tools 1, 2, 3 and 6 are new, and the other two are modified versions of well known tools. The new tools have been described in Pickup et al and Wilson et al in this volume and so no additional information is given here regarding their derivation or development.

2.1 Operational Demand Evaluation Checklist (ODEC)

ODEC has been administered in all instances of the workload assessments. The checklist collects a set of data that reflects the ‘objective’ demand on the signaller, for example number of track miles controlled. The assessors classify each answer into a high, medium or low category; a simple scoring system is used (‘3, 2, 1’ for high, medium or low respectively). Some high salience items in the checklist have been identified as scoring double.

Although the data can be collected from a number of sources, the assessors try to access the data in the way that is least disruptive to the signaller. Data are collected from the signal box specific Sectional Appendices (eg track miles and line speed), the signalling display (controlled and automatic signals), and the simplifier (timetable/number and frequency of planned traffic). If there is a shift manager, they can provide some of the answers that are not directly available to the assessors (eg short notice traffic movements and number of incidents in the last week). In the beginning, some of the categorisation of the data into high, medium or low had to be carried out by the operational specialist(s) based on general experience. This issue is addressed in a later section.

The valuable output of ODEC is the list of high scoring ‘demand’ items; these are useful indications of the demand factors that may underlie the workload ‘problem’. An overall summed ODEC score can be obtained. This score in isolation is of limited value, but when signal boxes are compared it is interesting to note the ordering of the total scores.

Results and revision of the ODEC tool

The experience of using ODEC has lead to a modified form of the checklist being developed by the assessors. The modifications and the reasons for them are as follows:

- The length of the checklist has reduced from 31 to 26 items. Items that have not helped discriminate between locations or appeared to have less relevance have been removed (eg tunnels and bridges).
- Ranges of scores for an item have been added, for example the number of points traversed by a simple and a complex train movement.
- A fourth category of ‘not applicable’ has been added for the assessments, because at some sites there are none of that item. This helps to distinguish between sites that have low and no instances of the item.
- Notes and prompts are provided to indicate the type of data required. It was noted that when the data from a number of assessments were compared, there
was a difference in interpretation of some of the entities to be assessed; eg for sidings, sometimes the data provided were the number of sidings whereas what is required is train movements in and out per day.

- A comprehensive scoring system has been developed for the checklist items. As we gathered more data, it became easier to be categorical about what constitutes high, medium and low for an item.
- Automatic scoring of the items in the spreadsheet has now been provided. Since we were entering the data into a spreadsheet, and because the scoring scheme was now defined, it was easy to construct the on-line version of ODEC.

Experience has made us concerned about including an overall score in the feedback reports. Its presence encourages some readers to attach too much weight to it as an absolute indication of the workload level.

This is evident when the scores from a mechanical lever frame are compared with a NX panel. The NX panel scores much higher (for instance, 96 versus 49). If these scores are taken at face value, a false impression of demand is given. There are genuine reasons for the differences: lever frames control small sections of route, have fewer signals, and many of the ODEC items do not exist at these sites (eg Hot Axle Box Detectors). In fact, the mechanical signal box that scored 49 was deemed to be a demanding signalling work position. The implication of this is that there needs to be a separate scoring system for mechanical lever frame sites. It is likely to be based on a multiplier scoring system. It is clear that we are not in a position to specify a ‘red line’ level based on the overall ODEC score.

Finally, a feature of ODEC that was not at first apparent to the assessors is that signallers can construe it to be a Job Evaluation method. Some signallers asked directly, ‘Will this lead to a re-grading of the job?’ It is important that the signallers are briefed at the outset to establish that this is not about job evaluation, however based on some of the outcomes and recommendations of these assessments it is not possible to say that the two are unrelated.

2.2 IWS and Timeline Analysis

These two tools are reported together to reflect that in the field they are administered concurrently. The timeline and IWS are typically recorded for a 60-minute period that coincides with the predicted peak workload. The results from the two tools allow the assessors to see a ‘trace’ of increases and decreases in reported levels of demand and effort and correlate with the activities represented in the timeline.

It became usual for the operational specialist to record the timeline information manually. A recording sheet divided into minutes was used to record the activity and give some indication of start and end if it ran to more than 30 seconds. It was usual for this duration to be exceeded for telephone calls, but less usual for any other signalling task eg route setting or data entry for Train Description. The timeline also recorded walking around the panel and watching a particular event. The activity was video recorded, but on no occasion have we needed to refer to it to amplify the timeline record.
The IWS was prompted every 5-minutes. Simulator based trials (see Mitchell in this volume) were undertaken with shorter inter-stimulus intervals, but in the live signalling situation a 5-minute interval appeared to give the correct trade-off between number of data points and interrupting the signaller to ask for a rating.

Two methods of capturing the IWS data were used. Initially, after briefing the signaller, the assessors set up the response options sheet next to the signaller’s work area and then prompted (every five minutes) for a verbal rating. The IWS was designed with 9 semantic levels of demand/effort without a numerical rating scale. For reasons of ease of use, we added a numerical value for each of the points.

Another method used was a customised watch (Actiwatch Score) that can be programmed to prompt at the specified intervals and record the rating that the signaller makes (by pressing a button several times until the appropriate number is shown). This method also worked well and has the advantage that the signaller can respond on the move. Here the scale was inverted so that ratings at times of highest load required less button presses from the signaller.

**Results**

The results gained show that the IWS ratings vary over the recording period and give a good indication of the relative demand and effort relating to the tasks recorded in the timeline.

The respondents appear to adopt a response set. For the initial series of ratings, the signallers refer to the response descriptions, but once they settle into a pattern of rating they appear to vary about their chosen ‘default’ level without referring to the scale description.

Prior to using the IWS there was concern that the 9-point scale may be unwieldy. In practice, it does not appear to be a problem. All levels of the scale have been recorded at different times (but not by the same person). It may be that the ‘response set’ effect described above makes the scale practical and easy to use.

**2.3 Workload Principles, SWAT, Rail Adapted Modified Cooper Harper (RMCH) and Probes (decision ladders)**

The Principles, SWAT and RMCH are short tools that can be used to address whether a workload assessment is necessary, or else used at the initial stage of the assessment to determine whether workload may be a problem at the outset.

The Probes (in the form of decision ladders) comprise an in-depth, diagnostic tool that surveys a broad area comprising Receiving Information, Giving Information and Evaluating and Decision Making. When a problem is detected at a general level, the tool is used to drill down with a sequence of related, detailed questions.
Results

These analytical tools have not always been administered during the assessment. There are two main reasons for this:

1. As explained earlier, the conduct of the assessment works best when signaller intervention is minimised. These tools rely on face-to-face interviewing of the signaller and require varying amounts of time. If the signal box is staffed by a single signaller, then it is difficult to administer. If there are a number of signallers or a relief signaller is present, then it becomes easier to administer these tools.

2. As in life, when visiting a signal box there comes a time when to remain is to out-stay your welcome. After concluding the ODEC, IWS and Timeline activities the assessors usually have been on site for 2 hours. At this point in the assessment, the assessors begin to drill down into areas that have emerged as the particular set of demands that may be related to workload. This activity is interspersed with statements that summarise the understanding of the assessors that the signaller is invited to confirm. This phase seems to be critical for the assessors to establish that they have detected the sources of workload and that they have an understanding of their effects.

On a number of occasions, the assessors have used impressions and perceptions gained during the assessment to complete some of these tools offline after leaving the site. The value added by this is not compelling, so it is now rarely completed in this way.

Within the 12 Principles, five are primary and are essential attributes of the signalling task in that compromising these is likely to lead to an erosion of safety and performance. On occasions these 5 primary principles alone have been administered to signallers to provide a pass/fail view of the signalling task. However, it is difficult to arrive at a categorical Yes/No response. The answers recorded will be different if one prefaces the questions with either:

- ‘In the main …’, or
- ‘In every instance …’

SWAT and RMCH have been least used of the tools. In addition to the reasons that have been explained above, the sequence in which an assessment takes place means that short, summary tools such as these seem out of place at the end of an in depth assessment. ODEC is the preferred starting point for the assessment since the data gathering is non-threatening and allows the signallers to become ‘comfortable’ with the assessors; it also enables the assessors to familiarise themselves with the signalling site characteristics. So, those tools that are directed towards detecting, at an early stage, whether there is workload problem, are somewhat redundant after a more in-depth assessment has already taken place.
The Probes have been administered on one occasion only. The tool comprises three parts and part 1 – Receiving Information required 60-minutes to administer. Part 2 – Giving Information was conducted with a second signaller and was completed in 15 minutes. Part 3 – Evaluating and Deciding was less well developed and was curtailed after 10-minutes.

It was clear that the Probes were effective in identifying the major sources of demand, understanding alternative strategies for addressing the demand, and establishing the safety, performance and signaller well-being consequences. Ranking of the factors was not performed on this occasion.

A continuation assessment visit is required when, after the first visit, the assessors have not identified whether or not there is a workload problem, or if there is a workload problem but they have failed to identify the contributing factors. On account of its extended time requirement and the need for a captive signaller for up to 2-hours, it was decided that the Probes would be best used when a continuation assessment visit was required.

The experience of administering this tool has established that it is a worthwhile, in-depth and diagnostic tool. A re-worked and re-packaged version is being developed.

3 Detection of Workload and Sources of Demand

In this section, the performance of the tools is evaluated and discussed with respect to:

1. detecting workload, and

2. diagnosis of the sources of task demands that cause workload problems for signallers.

Based on the results gained from the initial assessments it cannot be claimed that the tools applied categorically indicate the presence of undesirable levels of workload. What the tools do provide is a standard and signaller relevant set of data that the assessors must weigh to decide whether there is a workload problem.

In particular, ODEC classifies the physical infrastructure elements of the signalling control area of responsibility into high, medium or low. A score is derived from the tool but, for reasons already expounded, it is inadvisable to consider this score as a categorical level of demand.

Similarly, the IWS scores indicates a rating of the load experienced and effort required by the signaller, but caution is needed in taking the average, minimum and maximum values as absolute evidence of workload demand. The IWS and the Timeline are most useful together in associating the activity being undertaken and the ratings of effort over a period of time.

So if these results are indicative rather than categorical, then what is the process that the assessors use to determine an answer to the two basic questions?
As already stated the assessors have not used the summary tools (SWAT, Principles and RMCH) as the starting point for the assessment. Instead, the order in tool application is normally ODEC then IWS and Timeline. Towards the end of the assessment, the assessors are able to draw on this detailed information to address the detection of workload question. To make the judgement, they need to be confident that they have understood the level of workload and sources of demand on the signaller. If they are confident, then the workload judgement is easy to make; if not, the workload assessment needs to be adjourned and further examination undertaken on a later continuation visit.

At the outset, the assessors had a concern that the workload assessment toolset might not discriminate between high and excessive workload, and would thereby indicate high workload in all sites visited.

Because of the method of referral for assessment, nearly all sites have been found to have high workload. In more than half of these sites, the reasons attributed for elevated levels of workload have been short-term deficiencies either created by the infrastructure or by the signaller’s equipment. In these cases, the assessors consider that once the deficiencies are addressed and remedied, the level of workload will return to an acceptable level. In less than half the assessments the recommendation for extra staffing resource has been made during particular periods where the demand on the signaller is particularly high. Re-grading of the job has not been recommended on any of these occasions.

4 Representativeness of the Results

For these assessments, there are a number of concerns about whether the workload measure is:

- representative of all the individual signallers and the range of task demands that confront them
- manipulated by the respondents
- restricted in generality due to the short period of time that the assessments were conducted.

Each of the assessments carried out have addressed and observed a single signaller’s experience of workload and are not necessarily a consensus view of the signalling team. On account of this, on occasions when we have suspected that the signaller studied is less representative of the rest of the staff, we have followed up the visit to meet other staff and validate the findings with these signallers’ perceptions. An instance of this was observed in a signal box where a highly experienced signaller participated in the assessment, but the other staff were relatively new to signalling.

The assessors are aware of the possibility that some manipulation of results might take place, especially when the staff are already specifically complaining about workload. This risk has been reduced by having an operational specialist on the assessment team. Also, the bias may be in the opposite direction. A few signallers have commented, after completing the IWS session, that they find it difficult to give
ratings at the higher levels of effort and that it ‘goes against the grain’. Therefore, there is a risk that signallers will downplay their perceived effort.

The timing and duration of the assessments is potentially troublesome. Since the assessment rarely exceeds 3 hours, there may be some doubt about the generality of the results based on this short window. However in practice, the remit of the assessment often specifically directs us to consider a particular period, for example the morning or evening rush, the periods of possession management, the combination of two areas of control for the night shift, and requests for line crossing by the public (eg ramblers, holiday makers). In the Workload Assessment Report, the prevailing state of the railway during the assessment is always described, and it is emphasised that the conclusions and recommendations are made on the basis of what was observed and recorded on the day. This provides a useful platform to consider variations in demand on the overall workload.

In summary, the three points of concern raised have, in practice, not had a significant effect on the generality of the results.

5 Concluding remarks

This paper presents the experience of applying a set of tools for workload assessment. The tools have been developed with the rail industry in mind and target the signalling task in particular.

Criteria for assessing the industrial efficacy of the tools have been stated in this paper. The conclusion, based on the assessments conducted to date, is that the tools generate the necessary information for the assessors to determine whether there is a workload problem, they are effective in identifying the sources of demand, and they can be administered within a half day.

Acknowledgements

This paper is based on the application of the toolset in a number of the signalling locations assessed. The assessors that collected these reported data were part of the Ergonomics Group at Network Rail and I gratefully acknowledge their support in preparing this paper. Finally, thanks are due to the signallers who agreed to participate in and were supportive of these assessments.