Reliability of Manual handling Assessment Charts (MAC) developed for health and safety inspectors in the UK

A field study

Sarah E Tapley MSc

Field Operations Directorate, HSE

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

OBJECTIVES

A project by the Health and Safety Executive (HSE) and the Health and Safety Laboratory (HSL) has resulted in the development of a tool to aid inspectors of health and safety in the UK when assessing the risks of manual handling operations in workplaces. This is known as the "Manual handling Assessment Charts", or MAC.

This study was designed to assess the reliability of risk assessments made using the draft MAC tool released in May 2002 for evaluation by HSE inspectors. A revised version was formally released to all HSE inspectors in November 2002.

MAIN FINDINGS

A series of inter and intra-rater reliability assessment exercises were carried out. Approximately 10% of HSE field inspectors carried out a survey exercise to score the risk factors in four different tasks using the MAC tool and recording their observations. Levels of agreement within groups were calculated out and comparisons made between groups to assess if there were differences in scoring by experts and non-experts, inspectors who had received a briefing about the risk factors and those that had not. Intra-rater reliability was assessed by using a test-retest approach.

The levels of agreement for all samples indicated strong to good association and there were few statistically significant differences in scoring between any of the sample groups. Those differences noted would appear to be due to the simulation methodology used during the testing process.

CONCLUSIONS

This tool would therefore appear to enable regulatory inspectors to reliably make an initial identification, assessment and evaluation of the significant the risk factors for manual handling tasks.

1 BACKGROUND

There is a plethora of methods available for use by ergonomists and practitioners when assessing exposure to risks associated with work-related musculoskeletal disorders (WRMSDs) (Li and Buckle, 1999b). However, many such methods have been developed for a particular research purpose, often based on the expert's view of which occupational risk factors should be considered and how they should be measured, resulting in methods that are sometimes so sophisticated that only researchers or well-trained analysts are able to use them (Li and Buckle, 1999a). This fact has been commented on by several other authors (Stanton and Young, 1997; Burt and Punnett, 1999; and Kemmlert, 1995).

There is little evidence in the literature of the reliability or validity of ergonomic methods. This was confirmed in a survey of professional ergonomists to evaluate their practices in the use of different ergonomic methods (Stanton and Young, 1997). This lack of evidence may be due to a lack of motivation on behalf of the people developing and using the methods to ensure they meet the development criteria set (Li and Buckle, 1999a). It has been suggested that reliability and validity may be application specific (Wilson, 1995). Stanton and Young (1997) have noted that an increased demand for inventive approaches to assess users and their requirements in the design of work seems to have led to the pragmatic development of methods having priority over scientific rigour. Li and Buckle (1999b) comment that tools may be well validated and be shown to be reliable for the context that they were developed for. However, if the tool is then taken and used in a different context, its reliability and validity may be questioned.

Several recent studies (Denis *et al.*, 2002; Lansdown *et al.*, 1994; Jones *et al.*, 1999) have found that training, experience and practice did not significantly impact on observer ability when making initial assessment of manual handling risk factors.

A project by the Health and Safety Executive (HSE) and the Health and Safety Laboratory (HSL) has resulted in the development of a tool to aid inspectors of health and safety in the UK when assessing the risks of manual handling operations in workplaces. This is known as the "Manual handling Assessment Charts", or MAC.

This project found that the potential for use of an existing tool in the inspection setting was restricted through failure of any one tool to possess all the criteria specified by the project team. As a result, a draft set of tools was produced for assessing lifting, carrying and team handling operations. Each chart used a flowchart format with a "traffic-light" risk indication system. This was initially released for evaluation by selected HSE inspectors in May 2001.

This study was designed to assess the reliability of risk assessments made using the draft of the MAC tool released across HSE in May 2002. A revised version of the MAC, taking suggestions from users and the recommendations from the evaluation work into account, was formally released to all HSE inspectors in November 2002 (HSE, 2002). The development of the tool has been described by Monnington *et al.* (2002). Usability testing is reported by Care *et al.* (2002) and benchmarking of the charts against other tools is reported by Pinder (2002).

2 METHODS

2.1 AIM

To assess whether using the MAC tool enables UK inspectors of health and safety to reliably identify the physical risk factors for manual handling injuries

2.2 OBJECTIVES

1. To assess levels of agreement in scoring risk factors by inspectors and experts

2. To assess if there is a difference in scoring risk factors between a group of ergonomists (experts) and a group of general inspectors (non experts)

3. To assess whether there is a difference in scoring risk factors between inspectors who received a briefing (briefed) before scoring and those who did not (non-briefed).

4. To assess the level of intra-rater scoring agreement in a sample of inspectors.

2.3 STUDY DESIGN

The study design used a quantitative survey methodology carried out in a controlled environment using conference room facility in HSE field offices. This was done as part of a mandatory briefing session for inspectors when the draft version of the MAC tool was being launched.

2.4 SAMPLE GROUPS

Sample groups were selected on a geographical convenience basis. Each sample group contained a mixture of grades of general health and safety inspectors who regularly dealt with manual handling issues as a part of their routine workload.

The sample groups were:

SG 1 Operational inspectors (n=50) in two HSE field offices attending a briefing session about the tool were given detailed briefing about the risk factors addressed in the tool before completing the scoring exercise and were not classed as experts (briefed non-experts).

SG 2 Operational inspectors (n=22) in one field office attending a briefing session on the use of the tool who received briefing about the risk factors addressed in the tool after they had completed the scoring exercise (non-briefed non-experts).

SG 3 Ergonomists (n=5) with two or more years experience in manual handling assessment working in HSL. This group also received the briefing about risk factors in the tool before scoring the tasks (briefed experts).

SG 4 Operational inspectors (n=8) from SG 1 who attended a second briefing session, six months later, on the use of the tool and then rescored the same tasks using the tool (test-retest briefed non-expert).

2.5 SURVEY METHOD

Video footage taken by an ergonomist from HSL when accompanying HSE inspectors during workplace visits was assessed. Clips were selected that clearly demonstrated the risk factors in the three tools. The final selection included two clips for lifting as this was felt to be the most likely scenario that inspectors would encounter when inspecting, and one clip each for carrying and team handling operations.

The tasks selected were:

Task 1 Lifting operation: Handling creels of wire onto spindles for use in wheel making process. Creels weighed 20 kg and were handled at a rate of 250 per hour over the 8-hour shift.

Task 2 Lifting operation: Compost bale handling from a pallet to the edge of a hopper where it was split and the contents tipped into the hopper from where an auger fed it onto a potting line filling flower pots. The bales weighed 50 kg and three were handled in quick succession, 12 bales per day by one operator.

Task 3 Carrying operation: Carrying tubs of pastry up steps to load cookers. The tubs weighed 20 kg and were carried between 4 and 10 metres at a rate of four carries per hour.

Task 4 Team handling: A four man team lifting and turning mould lids used in vehicle trim and matting manufacture. The lids weighed 110 kg.

Information that would have been elicited by questioning and observation during a site visit was given in a verbal protocol at the start of the scoring exercise; this included the load weight and repetition rate, carrying distance, communication and coordination, floor surface and other environmental factors.

Inspectors were asked to score each task using the MAC tool and its associated score sheet. Each task was discussed after the score sheets had been collected, for training purposes, explaining what score the ergonomist who carried out the original work had reached, how that score had been obtained and what action had been taken as a result of the inspection and highlighting possible control measures.

2.6 DATA ANALYSIS

The collected data were analysed with SPSS v10 to assess the following:

- The level of agreement on scoring risk factors within each sample group for each task, computed using the Kendall coefficient of concordance.

- The difference in scoring between experts and non-experts, computed using the Mann-Whitney test.

- The difference in scoring between briefed and non-briefed inspectors, computed using the Mann-Whitney test.

- The difference in scoring by inspectors carrying out scoring (Test 1) and then repeating the scoring (Test 2) six months later, computed using the Wilcoxon test.

Statistical significance levels were set at the conventional level of $p<0.05$.

3 RESULTS

The results obtained from statistical testing carried out on data gathered during the survey are set out below:

3.1 SAMPLE GROUP DATA

Table 1 indicates that the subjects had a wide range of experience, from less than 1 year to 28 years, with a mean value of 8.35 years. SG 1 and SG 2 were a mix of grades of HSE staff involved in visiting workplaces to enforce and advise on health and safety legislation in the UK. SG 3 was a group of ergonomists from HSL who are regularly called upon by HSE field operations staff for specialist advice on manual handling. SG 4 was a subset of SG 1.

Table 1 Sample group data

Staff seniority	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 4 (n=8)***
Band 2	8	3	1 (Principal ergonomist)	0
3	30	11	1 (Senior ergonomist)	7
4	5	8	4 (Ergonomist)	0
5	2	0	0	1
Not known	5	0	0	0
Years of experience				
Range	0.25 - 25	0.25 - 28	2 - 20	2 - 12
Mean	8.57	7.63	8.2	9.00

3.2 LEVELS OF AGREEMENT

Levels of agreement in scoring all risk factors for each task for SG 1, SG 2 and SG 3 are summarised in Table 2. Results were computed using Kendall's coefficient of concordance (*w*). A *w* result may range from 0 (no association) to 1 (total association). A score of $w > 0.9$ indicates a strong association; a *w* score between 0.7 and 0.8 indicates a good association (MG Boocock, personal communication).

Table 2 Overall level of agreement within the different sample groups

Task	*SG 1 (n=50) Briefed non-experts*	*SG 2 (n=22) Non-briefed non-experts*	*SG 3 (n=5) Briefed experts*
1	*w* = 0.91	*w* = 0.93	*w* = 0.89
2	*w* = 0.93	*w* = 0.99	*w* = 0.77
3	*w* = 0.89	*w* = 0.97	*w* = 0.93
4	*w* = 0.78	*w* = 0.82	*w* = 0.77

w = strong agreement at >0.9
w = good agreement at >0.7

Assessment of the results indicates that there is a good to strong agreement between inspectors identifying the risk factors since all *w* values exceed 0.7 (the minimum observed was 0.77). It is not possible to make direct comparisons between groups, but, the generally similar ranges and distribution of scores for the groups add weight to this finding, allowing comment to be made that the level of agreement is good.

3.3 DIFFERENCE IN SCORING BETWEEN EXPERTS AND NON-EXPERTS

Table 3 Mann-Whitney *z* scores and corresponding *p* values for differences in scoring between experts (SG 3) and non-experts (SG 1)

Variable	*Task 1*	*Task 2*	*Task 3*	*Task 4*
Load weight	z = -0.90 p = 0.37	z = 0.00 p = 1.00	z = -0.46 p = 0.65	z = 0.00 p = 1.00
Hand distance from low back	z = -0.30 p = 0.77	z = -1.46 p = 0.15	x	z = -1.52 p = 0.25
Vertical lift distance	z = -0.56 p = 0.57	**z = -3.5** **p = 0.00**	x	z = -1.55 p = 0.12
Trunk twisting/asymmetry	z = -1.09 p = 0.27	z = -0.27 p = 0.79	Insufficient data	z = -1.01 p = 0.32
Postural constraints	z = -0.95 p = 0.35	z = -0.94 p = 0.35	z = -0.11 p = 0.92	z = -1.13 p = 0.26
Grip on load	z = -0.36 p = 0.72	z = -0.64 p = 0.53	z = -0.32 p = 0.75	z = -1.43 p = 0.15
Floor surface	z = -0.65 p = 0.52	z = -0.07 p = 0.95	z = -0.90 p = 0.37	z = 0.00 p = 1.00
Other environmental factors	z = -0.32 p = 0.75	z = -0.63 p = 0.53	z = -0.90 p = 0.37	z = -0.74 p = 0.46
Carry distance	x	x	z = -0.46 p = 0.65	x
Obstacles en route	x	x	z = -0.46 p = 0.65	x
Communication and coordination	x	x	x	**z = -2.5** **p = 0.01**

x indicates that the risk factor was not scored in that task
Values in bold show where p was significant at <0.05

Table 3 summarises the *z* scores and corresponding *p* values obtained when the differences in scores given by the experts and non-experts were analysed using the Mann-Whitney test.

Assessment of the results would indicate that there is no statistically significant difference in scoring for all but two of the risk factors. These two factors were Vertical lift distance in Task 2 and Communication and coordination in Task 4. Information for this last factor was supplied at the start of the task scoring and so it would have been reasonable to expect no difference in scoring at all.

3.4 DIFFERENCE IN SCORING BETWEEN BRIEFED AND NON-BRIEFED INSPECTORS

Table 4 summarises the *z* scores and corresponding *p* values obtained when the differences in scores from briefed and non-briefed inspectors were analysed using the Mann-Whitney test. Assessment of the results would indicate that there is no statistically significant difference in scoring for all but five of the risk factors. Statistically significant differences were found for Hand distance from low back and Vertical lift distance in Task 1 and Grip on load in Task 2. Differences in Task 4 were shown for Vertical lift distance and, as in the previous results, Communications and coordination.

Table 4 Mann-Whitney *z* scores and corresponding *p* values for differences in scoring between briefed (SG 1) and non-briefed (SG 2) inspectors

Variable	*Task 1*	*Task 2*	*Task 3*	*Task 4*
Load weight	z = -0.07 p = 0.95	z = 0.00 p = 1.00	z = -0.95 p = 0.34	z = 0.00 p =1.00
Hand distance from low back	**z = -2.53** **p = 0.01**	z = -0.50 p = 0.62	x	z = -0.36 p = 0.70
Vertical lift distance	**z = -2.18** **p = 0.03**	z = -0.68 p = 0.50	x	**z = -2.02** **p = 0.04**
Trunk twisting/asymmetry	z = -0.67 p = 0.51	z = -1.24 p = 0.22	z = - 1.68 p = 0.09	z = -1.32 p = 0.19
Postural constraints	z = -0.85 p = 0.40	z = -1.20 p = 0.23	x	z = -0.45 p = 0.63
Grip on load	z = -0.27 p = 0.79	**z = -2.50** **p = 0.01**	z = -1.28 p = 0.20	z = -0.92 p = 0.36
Floor surface	z = -0.53 p = 0.60	z = -0.11 p = 0.91	z = -0.67 p = 0.50	z = 0.00 p = 1.00
Other environmental factors	z = -0.66 p = 0.51	z = -0.98 p = 0.33	z = -1.85 p = 0.06	z = -0.15 p = 0.88
Carry distance	x	x	z = -0.95 p = 0.34	x
Obstacles en route	x	x	z = -1.45 p = 0.15	x
Communication and coordination	x	x	x	**z = -3.18** **p = 0.001**

x indicates that the risk factor was not scored in that task
Values in bold show where p was significant at <0.05

3.5 SCORING OF FACTORS IN THE FLOWCHARTS

Each risk factor has a Red, Amber or Green scoring choice. The following tables indicate the frequencies, as percentages, of these choices for each task by each group. Results were computed using the descriptive statistics function in SPSS v10. These results indicate the range of scoring as well as the frequencies and demonstrate the similarities in ranges and frequencies among the groups.

3.5.1 Load weight factor

This factor was scored by reading the load/frequency graphs, which are integral to both the lifting and carrying flowcharts in the MAC tool, or by using the weight categories in the team handling tool. Table 5 indicates that the frequencies were almost identical for all groups for this factor.

Table 5 Scoring of the Load weight factor for the four tasks

	Task 1			***Task 2***		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Purple	0%	0%	0%	100%	100%	100%
Red	77.8%	77.3%	60%	0%	0%	0%
Amber	22.2%	22.2%	40%	0%	0%	0%
Green	0%	0%	0%	0%	0%	0%
	Task 3			***Task 4***		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Purple	0%	0%	0%	0%	0%	0%
Red	0%	0%	0%	98%	100%	100%
Amber	94%	100%	100%	2%	0%	0%
Green	4%	0%	0%	0%	0%	0%

Values for some tasks do not add to 100% because of missing data

3.5.2 Hand distance from low back

Table 6 indicates that the scoring of this factor for Task 4 is almost identical in all groups. Briefed non-experts and experts scored similar frequencies for Task 1 but the non-briefed group had a higher Red frequency than the other groups.

The inspectors' scores (SG 1 and SG 2) were very similar for Task 2 but the experts (SG 3) had a wider range across the factors. Task 3, the carrying operation does not include this risk factor. Scores were almost identical for Task 4.

Table 6 Scoring of the Hand distance from low back factor on the lifting / team handling tasks

	Task 1			*Task 2*			*Task 4*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	28%	59.1%	20%	64%	72.7%	20%	98%	100%	100%
Amber	70%	40.9%	80%	32%	27.3%	40%	2%	0%	0%
Green	2%	0%	0%	0%	0%	40%	0%	0%	0%

Values for some tasks do not add to 100% because of missing data

3.5.3 Vertical lift distance

Table 7 indicates that the two groups of inspectors gave similar frequencies of Amber scores for Task 1 but other scoring was varied for this factor. For Task 2 there was almost no variation between inspectors but a wider range by the experts. Task 4 indicates that non-briefed inspectors and the experts have a similar scoring frequency scoring of predominantly Green, but the briefed, non-experts scored Amber more frequently than Green.

Table 7 Scoring of the Vertical lift distance on the lifting / team handling tasks

	Task 1			*Task 2*			*Task 4*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	28%	9.1%	40%	94%	100%	60%	2%	4.8%	0%
Amber	66%	72.7%	20%	2%	0%	20%	70%	38.1%	40%
Green	6%	18.2%	40%	0%	0%	20%	26%	57.1%	60%

Values for some tasks do not add to 100% because of missing data

3.5.4 Trunk twist/asymmetry

Table 8 indicates that the scoring of these factors was broadly similar for all the groups.

Table 8 Scoring for the Trunk twist/asymmetry factors for the four tasks

	Task 1 (Trunk twisting / sideways bending)			*Task 2 (Trunk twisting / sideways bending)*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	80%	86.4%	100%	28%	13.6%	20%
Amber	18%	13.6%	0%	48%	59.1%	60%
Green	2%	0%	0%	20%	27.3%	20%
	Task 3 (Asymmetrical trunk/load)			***Task 4 (Trunk twisting / sideways bending)***		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	14%	0%	0%	4%	0%	0%
Amber	74%	100%	100%	38%	28.6%	20%
Green	4%	0%	0%	56%	71.4%	80%

Values for some tasks do not add to 100% because of missing data

3.5.5 Postural constraints

Table 9 indicates that the frequency of Amber scores in Task 1 was very similar, but there is a range for the other scores. The scores for Task 2 are almost identical for experts and non-briefed inspectors but the briefed, non-expert sample has a wider range of scores. The scoring by all the groups was very similar for Task 4.

Table 9 Scoring of the Postural constraints factor for the four tasks

	Task 1			*Task 2*			*Task 4*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	10%	0%	20%	4%	0%	0%	0%	0%	0%
Amber	80%	90.9%	80%	48%	72.7%	80%	18%	14.3%	40%
Green	10%	9.1%	0%	44%	27.3%	20%	80%	85.7%	60%

Values for some tasks do not add to 100% because of missing data

3.5.6 Grip on load

Table 10 indicates that the scoring for Task 1 was broadly similar. Frequencies of scores for Task 2 were similar for briefed, non-expert and experts; the non-briefed sample scored Amber more frequently than the other groups. Task 3 indicates that Inspectors scored Red more frequently than the expert group who scored Amber more frequently. For Task 4, Experts scored Amber more frequently than the inspector groups who scored predominantly Green.

Table 10 Scoring of the Grip on load factor for the four tasks

	Task 1			*Task 2*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	0%	4.5%	0%	64%	31.8%	80%
Amber	68%	63.6%	60%	28%	68.2%	20%
Green	32%	31.8%	40%	4%	0%	0%
	Task 3			***Task 4***		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	78%	95.4%	20%	0%	0%	0%
Amber	20%	4.5%	60%	28%	19%	60%
Green	0%	0%	20%	70%	81%	40%

Values for some tasks do not add to 100% because of missing data

3.5.7 Floor surface

Table 11 indicates that there was almost identical scoring of this factor by the different subject groups.

Table 11 Scoring of the Floor surface factor for the four tasks

	Task 1			*Task 2*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	0%	0%	0%	0%	9.1%	0%
Amber	8%	4.5%	0%	18%	9.1%	20%
Green	92%	95.5%	100%	78%	81.8%	80%
	Task 3			***Task 4***		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	98%	100%	100%	0%	0%	0%
Amber	2%	0%	0%	0%	19%	0%
Green	0%	0%	0%	98%	81%	100%

Values for some tasks do not add to 100% because of missing data

3.5.8 Carry distance and obstacles en-route

This factor was present only in Task 3. Table 12 indicates that there was almost the same scoring frequency for each group.

Table 12 Scoring of the Carry distance and Obstacles en route factors in Task 3

	Carry distance			*Obstacles en route*		
	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***	***SG 1 (n=50)***	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	4%	0%	0%	4%	13.6%	0%
Amber	94%	100%	100%	94%	86.4%	100%
Green	0%	0%	0%	0%	0%	0%

Values for some tasks do not add to 100% because of missing data

3.5.9 Communication and coordination

Table 13 indicates that non-briefed inspectors and experts scored Communication and coordination with a predominance of Amber scores. However, the briefed, non-expert sample scored Amber and Green more equally.

Table 13 Scoring for Communication and coordination in Task 4

	SG 1 (n=50)	***SG 2 (n=22)***	***SG 3 (n=5)***
Red	0%	0%	0%
Amber	40%	81%	100%
Green	58%	19%	0%

Values for some tasks do not add to 100% because of missing data

3.6 TEST/RETEST DIFFERENCES BETWEEN TEST 1 AND TEST 2 SCORING

SG 4 was a subset of SG 1. They therefore received the briefing about the risk factors before carrying out the scoring exercise. Six months later they were given a fresh briefing and carried out the scoring exercise again. Assessment of the results, in Table 14 shows there was no statistically significant difference in scoring between Test 1 and Test 2 for any of the tasks. The only close result was for Communication on Task 4.

Each risk factor has a Red, Amber or Green scoring choice. The following sections indicate the distribution of scores for each task by each group. These results indicate the range of scoring as well as the frequency and demonstrate the similarities in ranges and frequencies between Test 1 and Test 2. This adds weight to the argument that the tool is reliable under test/retest assessment.

Table 14 Mann-Whitney z scores and corresponding p values for test-retest differences in scoring

Variable	*Task 1*	*Task 2*	*Task 3*	*Task 4*
Load weight	z = -1.00 p = 0.32	z = 0.00 p = 1.00	z = 0.00 p = 1.00	z = 0.00 p = 1.00
Hand distance from lower back	z = -1.00 p = 0.32	z = -1.00 p = 0.32	x	z = -1.63 p = 0.10
Vertical lift distance	z = -1.41 p = 0.16	z = 0.00 p = 1.00	x	z = -0.58 p = 0.56
Trunk twisting/asymmetry	z = 0.00 p = 1.00	z = -0.58 p = 0.56	z = 0.00 p = 1.00	z = -1.34 p = 0.18
Postural constraints	z = -1.41 p = 0.16	z = -0.18 p = 0.85	Insufficient data	z = -1.41 p = 0.16
Grip on load	z = -1.00 p = 0.32	z = -1.73 p = 0.08	z = -0.58 p = 0.56	z = -1.41 p = 0.16
Floor surface	z = -1.00 p = 0.32	z = -1.00 p = 0.32	z = 0.00 p = 1.00	z = 0.00 p = 1.00
Other environmental factors	z = -1.00 p = 0.32	z = -1.00 p = 0.32	z = -0.58 p = 0.56	z = -0.58 p = 0.56
Carry distance	x	x	z = 0.00 p = 1.00	x
Obstacles en route	x	x	z = -1.00 p = 0.32	x
Communication and coordination	x	x	x	**z = -2.00** **p = 0.05**

x indicates that the risk factor was not present in that task
Values in bold show where p was significant at <0.05

3.6.1 Task 1

Assessment of the scoring distributions would indicate that there is a consistent range of scoring for most of the risk factors, with the exception of vertical lift distance where there is a wider range for Test 1 than Test 2.

Table 15 Test-retest scores for Task 1

	Load weight		*Hand distance from lower back*		*Vertical lift distance*		*Trunk twisting/ asymmetry*	
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***
Purple	0%	0%	0%	0%	0%	0%	0%	0%
Red	87.5%	62.5%	0%	0%	12.5%	0%	87.5%	87.5%
Amber	12.5%	37.5%	100%	87.5%	75%	50%	12.5%	12.5%
Green	0%	0%	0%	12.5%	12.5%	50%	0%	0%
	Postural constraints		***Grip on load***		***Floor surface***		***Other environmental factors***	
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***
Purple	0%	0%	0%	0%	0%	0%	0%	0%
Red	0%	0%	0%	0%	0%	0%	0%	0%
Amber	100%	75%	50%	75%	0%	12.5%	0%	12.5%
Green	0%	25%	50%	25%	100%	87.5%	100%	87.5%

3.6.2 Task 2

Assessment of Table 16 would indicate that the scoring of Task 2 was broadly similar for both the test and retest.

Table 16 Test-retest scores for Task 2

	Load weight		*Hand distance from lower back*		*Vertical lift distance*		*Trunk twisting/ asymmetry*	
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***
Purple	100%	100%	0%	0%	0%	0%	0%	0%
Red	0%	0%	87.5%	100%	100%	100%	12.5%	12.5%
Amber	0%	0%	12.5%	0%	0%	0%	75%	62.5%
Green	0%	0%	0%	0%	0%	0%	12.5%	25%
	Postural constraints		***Grip on load***		***Floor surface***		***Other environmental factors***	
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***
Purple	0%	0%	0%	0%	0%	0%	0%	0%
Red	12.5%	12.5%	62.5%	100%	0%	0%	0%	0%
Amber	50%	50%	37.5%	0%	0%	12.5%	62.5%	87.5%
Green	37.5%	37.5	0%	0%	100%	87.5%	37.5%	12.5%

3.6.3 Task 3

Assessment of Table 17 would indicate that the scoring of Task 3 was almost identical for both test situations.

Table 17 Test-retest scores for Task 3

	Load weight		***Carry distance***		***Asymmetrical trunk/load***		***Obstacles en route***	
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***
Purple	0%	0%	0%	0%	0%	0%	0%	0%
Red	0%	0%	0%	0%	0%	0%	0%	0%
Amber	100%	100%	0%	0%	0%	87.5%	100%	100%
Green	0%	0%	100%	100%	100%	12.5%	0%	0%
	Grip on load		***Floor surface***		***Other environ-mental factors***			
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***		
Purple	0%	0%	0%	0%	0%	0%		
Red	0%	0%	100%	100%	0%	0%		
Amber	87.5%	75%	0%	0%	25%	37.5%		
Green	12.5%	25%	0%	0%	75%	62.5%		

3.6.4 Task 4

Assessment of Table 18 would indicate that while there was no difference in scoring of Task 4 for the first two factors and the Floor surface factor, the remaining factors did show variation in the scores. The Vertical lift varied more in Test 2 than Test 1. The Trunk twisting factor had a wider range in Test 1 than in Test 2. The Communication and coordination factor was split between Green and Amber in Test 1 but was totally Amber for Test 2.

Table 18 Test-retest scores for Task 4

	Load weight		*Hand distance from low back*		*Vertical lift distance*		*Trunk twisting / sideways bending*		*Grip on load*	
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***
Purple	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Red	100%	100%	0%	0%	0%	12.5%	25%	0%	0%	0%
Amber	0%	0%	87.5%	87.5%	87.5%	50%	37.5%	50%	25%	0%
Green	0%	0%	12.5%	12.5%	12.5%	37.5%	37.5%	50%	75%	100%
	Communication and control		***Postural constraints***		***Floor surface***		***Other environ-mental factors***			
	Test 1	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***	***Test 1***	***Test 2***		
Purple	0%	0%	0%	0%	0%	0%	0%	0%		
Red	0%	0%	0%	0%	0%	0%	12.5%	25%		
Amber	50%	100%	37.5%	12.5%	100%	100%	87.5%	75%		
Green	50%	0%	62.5%	87.5%	0%	0%	0%	0%		

3.7 QUALITATIVE DATA

As discussed in the methodology section, the study was designed to collect quantitative data. During the scoring exercise some of the inspectors were quite vociferous in their comments. Verbal comments noted on several occasions in all the sessions included:

- 'We know all about manual handling.'
- 'I've been doing this job for years now, why do I need a checklist?'
- 'I know what the risk factors are.'
- 'What do we do with the scores we get.'
- 'I didn't realise grip was so important.'

Some of these comments may go some way to helping to explain some of the quantitative results obtained and will be discussed further in the next section.

4 DISCUSSION

The lack of differences in scoring between experts and non-experts and briefed and non-briefed inspectors is encouraging and is a similar finding to other recent reports (Lansdown *et al.*, 1994; Jones *et al.*, 1999; and Denis *et al.*, 2002), which have suggested that training and expertise do not affect the ability to make initial assessments of hazardous tasks. However, there are some discrepancies, as outlined in the previous sections, and there may be several reasons, such as the nature of the test material, for the discrepancies found.

4.1 FACTORS AFFECTING SCORING

There are good to strong levels of association in scoring across all the tasks and most of the risk factors with minimal statistically significant differences in scoring. These reasons, while not affecting the initial ability to identify hazardous factors, may have ramifications for the inspection process and future inspector training in manual handling and WRMSD issues.

4.1.1 Test material

The analysed data shows no difference in scoring for the majority of risk factors by all sample groups, which would indicate that all the observers were able to identify risk factors. This is consistent with other recent and similar studies (Denis *et al.*, 2002; Jones *et al.*, 1999; Lansdown, 1994).

However, there are some risk factors where there is less consistency in scoring, even though it may not be statistically significant and this may be due to a number of different factors. Denis *et al.* (2002) commented that one of the major factors affecting observer reliability could be the difficulty of using videotaped material. Troup and Rauhala (1987) suggested that use of videotapes can lead to loss of stereoscopic vision and a corresponding loss of scenario detail.

Qualitative comments indicated that the *Hand distance from low back* was the main risk factor that inspectors found difficult to visualise and hard to assess from video footage rather than being able to move around the task as they would in an inspection situation. This view appears to be supported as there is a significant difference between the briefed and non-briefed sample on Task 1 ($p = 0.01$).

Further qualitative comments were made about the *Grip on load* risk factor and that inspectors had not appreciated how important the grip on a load was during a manual handling operation. There was a significant difference between briefed and non-briefed inspectors for this factor in Task 2. The importance of grip may not have been adequately covered during the initial briefing about the risk factors addressed by the MAC.

Inspectors in the test/retest group commented during the retest scoring exercise that they had been using the MAC charts operationally during inspections since the first briefing. The main point they made was that they had found scoring the *Vertical lift distance* risk factor much easier when they were able to move around the task to view it from all angles and to be able to take reference points from other equipment in the surrounding areas.

The risk factor for *Postural constraints* in the tool prompts the user to consider the physical constraints on the worker adopting suitable postures to carry out a task. It is not requiring, as do other tools, such as the QEC (Li and Buckle, 1999a), an analysis of the posture adopted by the worker. This consideration can be difficult to assess from video footage when it is not possible to move around the task as one would in an inspection setting.

It can be seen that there are some drawbacks in using videotapes, however they do ensure that controlled testing conditions can be achieved with all scorers seeing the same task in the same way. The working group feedback indicated that the real life nature of the clips would compensate for lack of detail and engage inspectors more readily. It could also be suggested that there is very little alternative to videotapes for this type of exercise as this is the tool most often used in ergonomics studies. These limitations of using videotaping of tasks should be considered when designing any future training programmes. It may be possible to explore the use of virtual reality techniques to make any future training interactive.

4.1.2 Delivery of survey scoring exercise by researcher

This may have been a factor, especially when there were differences recorded for risk factors where information was given prior to the exercise, for example information about communication and coordination in Task 4. A verbal protocol was prepared to try to ensure that the same information was given in the same way to all the groups to ensure consistency of delivery. It could therefore have been assumed that there would be 100% agreement for all those risk factors but this was not the case.

The researcher may not have spoken clearly or loudly enough to ensure that everyone in the room clearly heard what was being said and possibly failed to check understanding after reading the protocol. This could account for statistically significant differences in scoring for vertical lift distance in Task 1 by briefed and non-briefed inspectors. Task 1 showed operatives loading creels of wire onto five different spindles from floor level to above shoulder height. The level at which the inspector decided to score would affect the scoring as the low and high spindles would attract a higher risk rating than the middle three spindles. A comment as to which spindles to score on during the briefing would have removed any confusion and may have made the scoring exercise more consistent. This demonstrates the importance of adequate briefing.

4.1.3 Potential divided attention

The sample groups may not have been paying full attention to the briefing; this may have been caused by the group having divided attention at the time of delivery of instructions. The inspectors looking at the charts, listening and reading the instructions at the same time could have caused this. Divided attention has the potential to cause performance decrements and may be caused during an inspection activity identifying manual handling risks by:

- Engaging in dialogue with one or a number of individuals;
- Trying to attend to simultaneous conversations during inspection;
- Listening and communicate with colleagues;
- Writing notes of a standard to be able to write a report at a later date;

- Recall of information both technical and legal to answer questions on the spot;
- Assessing the gap between industry standard and observed practice;
- Pursuing a consistent line of investigation through changing circumstances;
- Being required to look, assess and comment on other issues during the primary investigation.

The inspector may also be exposed to environmental stimuli such as heat, cold and plant machinery noise. The psychological capacity theory works with a very similar example of listening to a lecture and writing it down in shorthand. If the concept is difficult then it will require more attention to listen and consequently the notes will not be so well written. This is very similar to an inspector being unfamiliar with the identification of manual handling risk factors, using a new tool and having to take notes at the same time. Potential decrements will vary depending upon experience. An inspector who has had little training or experience in dealing with manual handling activities is likely to have higher decrements than an experienced inspector who is familiar with and experienced in dealing with the subject area. Decrements may include:

- Missing key points;
- Failing to record information, recording it incorrectly or incompletely;
- Misinterpreting information.

4.1.4 Potential inspection issues

The MAC tool is designed to avoid this potential performance decrement by leading inspectors logically through the risk factors. Levels of agreement and lack of difference in the majority of risk factor scoring would indicate that inspectors are identifying the factors reliably. Jones *et al.* (1999) found that all observers were able to identify hazardous tasks but could not easily prioritise them. As inspectors are being asked to identify and assess hazardous tasks and to rank them according to a colour banding but not to do any deeper assessment, this would suggest that the tool is fit for its purpose of assisting inspectors to identify significant manual handling tasks.

Variations in the ranges of percentage scoring for all those factors where information was given may also have been due to the underpinning knowledge, experience and attitudes of inspectors attending the briefing. Comments made at the start of the sessions indicated that inspectors knew that manual handling and manual handling injuries were an issue but many of the groups felt less confident about tackling them as they were less familiar with the so called ‘soft’ issues than with ‘harder’ issues such as machinery guarding. This related back to the difficulty in easily establishing the dose: response relationship for WRMSDs. These comments would appear to be corroborated by the low levels of enforcement action taken on manual handling issues by HSE inspectors. In the period from April 2001 to October 2002 only 1% of all enforcement notices and one prosecution were taken on manual handling issues.

Experienced inspectors will have a high degree of confidence in their ability, based on a wide range of health and safety experience, to go straight to the answer to a problem rather than having to work through a process to get to that answer. Some of the sample may have used

their prior knowledge and experience of the workplace to assume that they understood and knew what the answers were without checking their understanding of the briefing information first. If this were the case, then inspectors should ensure that they check and recheck all the information during an inspection to ensure all the information is correct - the logical approach of the MAC tool is designed to prevent this potential mistake.

4.1.5 Qualitative data

Verbal qualitative comments noted included 'I know what causes back pain' and 'we know all about manual handling'. These comments came predominantly from inspectors with more than 10 years' experience who were confident in their abilities as general health and safety inspectors. The range of scoring and differences in scoring some risk factors may be due to the training and experience that they have received to date. Ten years ago, workplace health and safety was predominantly safety orientated due to the lower level of knowledge about the cause and effect of health related issues and, in particular, the lack of accepted knowledge on WRMSD issues. This orientation during training may have influenced the inspector baseline mindset to understand and to believe that manual handling may be an issue but to not fully understand the extent and mechanics of the problem.

Other qualitative comments made were 'we know that there is a problem, but there is no point doing anything about it unless there is a solution'. The ranges of scoring and differences for some risk factors may indicate that, due to the lack of training to date on this issue, inspectors are as yet unable to break a task down into its component parts and look to see how each factor can be reduced, and so decrease the risk overall. This appears to agree with the findings of the study carried out by Jones *et al.* (1999), that observers are able to identify the hazardous tasks but not able to prioritise them and take action.

4.1.6 Session timing

The timing and duration of the survey session may too have had an impact on the scoring levels. The sessions were all run in the morning, this meant that the scoring exercise for Task 4 was done just before lunch at the end of the session; some of the lowest levels of agreement of scores were for Task 4. Accepted attention span is for 30 to 40 minutes before performance decrement occurs and there may also have been an element of fatigue and hunger just before lunch, which may have also decreased performance. Review of the briefing session would suggest that a break midway through the session would have been beneficial. This could have implications for the inspection process, if an inspector is to inspect whilst tired and hungry they may fail to gather all the information required. This is one of the arguments for the flowchart approach by following the charts in a logical and ordered manner it should be less likely for inspectors to miss any of the risk factors. This equates to training to a level where a person can operate on an automatic level, thus allowing spare processing capacity to deal with novel situations and so reduce the possibility of performance decrements.

4.1.7 Sampling factors

Ideal samples for any form of study would be randomly selected from the defined population using a sampling frame to try to reduce systematic bias. This study was largely dictated by operational and time constraints as to where and how the samples could be drawn. This resulted in the use of a geographical convenience sample within the same operating division as the researcher. However, the samples were purposive as they were drawn from operational HSE inspectors, which was the population of interest.

The final inspector sample size of 72 is approximately 10% of HSE's operational field inspectors who inspect and address the issue of manual handling on a regular basis. This sample size could therefore be said with some confidence as being representative and the findings extrapolated across all field inspectors. The expert sample (n=5) is approximately 25% of HSE ergonomists and so can also be said to be representative. The test/retest sample group (n=8) is small and so results should be taken with some caution. However, the frequency of scoring is very similar across the task to the other groups and this could be said to increase confidence in findings of the study.

4.2 TOOL DESIGN ISSUES

As stated previously, the detailed development of the tool is the topic of a separate report (Monnington *et al.*, 2002). However, some of the issues around the design of the tool have impacted on this study. An example is the *Load weight* risk factor. Information about the load/frequency was given before scoring was carried out using a verbal protocol to ensure consistency. The data show some differences in scoring, although the percentage frequency of scores is very similar across Tasks 2, 3 and 4. There was a wider range on Task 1, where the load weight score was borderline Amber/Red whereas the scoring for the other three tasks was almost 100%. Concerns about reading the scores from the graph had been raised during the tool development and comments were made by many observers during the scoring exercise. This was not an issue with the other tasks as the load weights were clearly in the relevant colour banding.

There were some discrepancies in data obtained for scoring of Task 3 that may have been due to the layout and wording of the score sheets. The wording on the score sheet ran in a different order to the MAC tool and no reference was made to *Asymmetrical trunk/load*. Qualitative comments received indicated that this had caused some confusion. The score sheet order also appeared to cause confusion for the *Trunk twisting/sideways bending* risk factor in the other tasks.

The purpose of the tool is to assist in identification and assessment of physical risk factors by regulatory inspectors. It could be argued, on the one hand, that there is growing evidence indicating a link between physical and psychosocial risk factors (Devereux *et al.*, 1999) and that this tool does not adequately address these. It could also be argued that, on the other hand, that assessment of psychosocial risks is far more complex than assessment of physical risk factors and judged to be not suitable for inclusion at this stage. However, use of the MAC tool will enable inspectors to become familiar with the physical risk factors to allow them, as discussed earlier, to operate at an automatic processing level. If and when the evidence on psychosocial factors is stronger and means of rapidly assessing them have been developed, it will be possible to incorporate these factors into a successor to the MAC.

4.2.1 Reliability issues

There are a variety of methods for assessing reliability, as set out in the review of the literature. This project used several methods to try to ensure that any findings were not purely down to chance agreement. The overall percentage of agreement using a correlation coefficient approach showed a good positive correlation in this study. However, Burt and Punnett (1999) commented that this was the most basic measure of reliability that could be open to a possibility of the findings occurring by chance. Fleiss (1973) suggests that use of the Kappa test takes into account the possibility of chance agreement. This test is suited to binary data

than can be organised into a contingency table. The data gathered in this study was not binary nor suited to a contingency table so the Kappa test was not suitable.

The reliability of this tool was further demonstrated by using an interrater approach where the same group tested and retested the tool acting as their own control group. The sample group was small (n=8) as discussed earlier, however the statistically significant lack of differences found in scoring risk factors adds weight to the argument that the tool is reliable.

The other method used to demonstrate reliability was interrater comparisons of expert against non-expert and briefed inspectors against non-briefed inspectors, which showed minimal differences in scoring, with good to strong levels of association. It could be argued that because the Kappa analysis was not used that the results of the testing were down to chance, however the variety of methods used which all record statistically significant results refute that argument.

4.2.2 Validity issues

The MAC tool was developed using validated research data and so can be said to demonstrate construct validity in the same way as the PLIBEL and QEC tools can. The tool is based on the accepted significant physical risk for manual handling injuries, however, it could be argued that the tool does not explicitly address other accepted risk factors such as vibration and the capability of the individual although they are recorded on the score sheet if present in a task. As Li and Buckle (1999b) point out, a tool may demonstrate very good reliability and validity for the task it has been designed for. However, if the tool is then taken and used in another context, the conclusions about its reliability and validity may not be so accurate. The MAC tool has been designed for use in identifying physical risk factors for manual handling injury. If the MAC was used to attempt to identify physical risk factors for Work Related Upper Limb Disorders (WRULDs), levels of reliability for scoring could be questioned as the full range or risk factors for WRULDs are not included in the MAC tool. It is important then, that any tool that is developed clearly sets out what it is intended for, to avoid a criticism of a lack of reliability and validity if it is used out of context.

5 CONCLUSIONS

5.1 GENERAL

The aim of the study was to assess whether inspectors could reliably identify the main physical risk factors for manual handling injuries using the MAC tool. A range of statistical tests was used for data analysis, and the results obtained would indicate that this is the case. It is therefore possible to say that all the objectives were fully met.

Levels of agreement found would indicate that inspectors are identifying the factors reliably. Jones *et al.* (1999) found that all observers were able to identify hazardous tasks, but could not easily prioritise them. As inspectors are being asked to make an initial identification and assessment of hazardous tasks and to rank them according to a colour banding rather than a detailed deeper assessment, this would indicate that the tool is fit for its purpose of assisting inspectors to identify significant manual handling risks.

The review of the literature identified that there is a plethora of assessment tools available for practitioners to use when assessing manual handling tasks; many of these tools, have been designed for specific activities or measuring exposure to specific parts of the body. However, it could be argued that to expect anything less would be against ergonomic principles as each assessment task is different, ergonomists seek constantly to dispel the myth that 'one size fits all' and this will surely apply to assessment tools as well.

The MAC tool, whilst designed for health and safety inspectors, would appear to be a potentially valuable addition to this battery of tests as an easy to use, reliable and valid initial assessment tool which can help practitioners to identify if and where further assessment is required.

5.2 SPECIFIC RECOMMENDATIONS

- Add grid lines onto the load/frequency graphs
- Revise the score sheet to ensure it runs in the same order with the same terminology as the flowcharts.
- Train inspectors in WRMSD issues to the extent that inspection of this topic becomes automatic, thus allowing them spare processing capacity to deal with novel situations.
- Provide further training for inspectors to make the link between psychosocial and physical risk factors
- Review charts as more evidence becomes available for the links and revise them accordingly

6 REFERENCES

1 Burt, S. and Punnett, L. (1999), Evaluation of inter-rater reliability for posture observations in a field study. *Applied Ergonomics*, **30**, (2), 121-135.

2 Care, B., Quarrie, C. and Monnington, S.C. (2002),*Testing and improving the usability of the Manual handling Assessment Chart (MAC)*. (Bootle: Health and Safety Executive). Internal report

3 Denis, D., Lortie, M. and Bruxelles, M. (2002) Impact of observers' experience and training on reliability of observations for a manual handling task. *Ergonomics*, **45**, (6), 441-454.

4 Devereux, J., Buckle, P.W. and Vlachonikolis, I.G. (1999), Interactions between physical and psychosocial risk factors at work increase the risk of back disorders: an epidemiological approach. *Occupational and Environmental Medicine*, **56,** (5), 343-353.

5 Fleiss, J. (1973),*Statistical methods for rates and proportions*. (New York: Wiley).

6 Health and Safety Executive (2002),*Manual Handling Assessment Charts*. (Bootle, Health and Safety Executive), MISC 480.

7 Jones, J.A.R., Cockcroft, A. and Richardson, B. (1999), The ability of non-ergonomists in the health care setting to make manual handling risk assessments and implementing change. *Applied Ergonomics*, **30,** (2) 159-166.

8 Kemmlert, K. (1995), A method assigned for the identification of ergonomic hazards - PLIBEL. *Applied Ergonomics* **26**, (3), 199-211.

9 Lansdown, T., Haslam, R.A. and Parsons, C.A. (1994), Development and evaluation of a manual handling assessment toolkit. In: Robertson, S.A. (ed) *Contemporary Ergonomics 1994*, 241-246.

10 Li, G. and Buckle, P. (1999a). *Evaluating change in exposure to risk for musculoskeletal disorders - a practical tool*. (Sudbury, Suffolk: HSE Books), HSE Contract Research Report no. 251/1999.

11 Li, G. and Buckle, P. (1999b), Current techniques for assessing physical exposure to work-related musculoskeletal risk, with emphasis on posture-based methods. *Ergonomics* **42**, (5), 674-695.

12 Monnington S.C., Pinder, A.D.J. and Quarrie, S.C. (2002), *Development of an inspection tool for manual handling risk assessment*. (Sheffield: Health and Safety Laboratory), HSL internal report ERG/02/20; HSL Report HSL/2002/30

13 Pinder, A.D.J. (2002),*Benchmarking of the Manual Handling assessment Charts (MAC)*. (Sheffield: Health and Safety Laboratory), HSL Internal Report ERG/02/21; HSL Report HSL/2002/31

14 Stanton, N. and Young, M. (1998), Is utility in the mind of the beholder? A study of ergonomic methods. *Applied Ergonomics*, **29**, (1), 41-54.

15 Troup, J.D.G. and Rauhala, H.H.(1987), Ergonomics and training. *International Journal of Nursing Studies* **24**, (4), 325-330.

16 Wilson, J.R. (1995), A framework and a context for ergonomics methodology. In: Wilson, J.R. and Corlett, E.N. (eds), *Evaluation of Human Work - A Practical Ergonomics Methodology*, (London: Taylor & Francis), Second Edition, 1-39.