Quality in fives

Oldenkamp, J.H.

Publication date:
1996

CHAPTER 6

AUDITING OF NURSING SCHEDULES

The previous chapter described the results of the ranking experiment. The conclusions of this ranking experiment consisted of five quality indicators. Each of these quality indicators measured the value of the corresponding factor of nursing schedule quality. On the basis of these conclusions, a second experiment was designed. The objective of this second experiment was to test the hypothesis of robustness, discussed in the third chapter. This hypothesis states that the total quality value of a nursing schedule according to a nurse scheduler can be explained on the basis of a weighted sum of the generic factor values.

The present chapter discusses this follow-up experiment. In this experiment, nurse schedulers were asked to audit several nursing schedules. This experiment is called the 'auditing experiment'. The following sections describe the design, results and conclusions of this auditing experiment.

6.1 DESIGN OF THE AUDITING EXPERIMENT

The objective of the auditing experiment is to test the hypothesis of robustness, which states that the total quality value of a nursing schedule s according to nurse scheduler i ($Q_{s,i}$) can be explained on the basis of a weighted sum of the values of the quality factors (i.e. completeness (C), optimality (O), proportionality (P), healthiness (H) and continuity (T)).

$$Q_{s,i} = \omega_i^{c} \times C_s + \omega_i^{o} \times O_s + \omega_i^{p} \times P_s + \omega_i^{h} \times H_s + \omega_i^{t} \times T_s$$

Formula 6.1: The audit value as a weighted sum of the factor values

To test this hypothesis, an auditing experiment was designed. In the design of this experiment, $Q_{s,i}$ was treated as the dependent variable. Furthermore, the factor values (C_s, O_s, P_s, H_s, and T_s) were designed to be the instrumental variables, and the summation weights (ω_i^c, ω_i^o, ω_i^p, ω_i^h, and ω_i^t) were designed to be the explanatory variables.

In this auditing experiment, five nurse schedulers were asked to give a quality mark on a scale from one to ten according to their own view on nursing schedule quality. These nurse schedulers were instructed to give unacceptable schedules a quality mark below 5.5, and acceptable ones a quality mark above 5.5.

The expected accuracy in giving these quality marks can be expressed as the standard error (σ) (see Lawson & Hanson, 1974). In this auditing experiment, the value of 0.5 is assumed to be realistic. This means that when a quality mark m is given to a nursing schedule by a nurse scheduler, the 'real' quality mark m^r of this nursing schedule for this nurse scheduler will have a value above m-2σ and below m+2σ in ninety-five percent of all cases. For example, when a mark of seven is given, there is a five percent chance that the really intended mark is below six or above eight.

Table 6.1 NUMBER OF ARRANGED NURSING SCHEDULES PER NUMBER OF UNACCEPTABLE FACTOR

UFV /sched	0	1	2	3	4	5	total
Psched / UFV	1	5	10	10	5	1	32
Asched / UFV	1	2	4	6	2	0	15

In total, the nurse schedulers were asked to give quality marks to fifteen nursing schedules for the fictitious nursing unit of East-5. For this nursing unit, the three independent variables — the length of the schedule period (N), the number of types of shift (D) and the staff size (M) — have the values of twenty-eight, three and twenty-three, respectively. Of these twenty-three nurses, nine are registered nurses and fourteen are nursing aids.

These fifteen fictitious nursing schedules differed in the number of quality factors with an unacceptable value, which is a value below 0.55. In total, thirty-six unacceptable factor values were distributed equally over the five quality factors. With an average of 7.2 times an unacceptable value per quality factor, four of the five quality factors have an unacceptable value in seven of the fifteen nursing schedules, while the remaining one — optimality — has an unacceptable value in eight of the fifteen nursing schedules. This equal distribution is important because the hypothesis of robustness, discussed in the third chapter, assumes that nurse

Table 6.2 INDICATION OF THE FACTOR VALUES OF THE FIFTEEN FICTITIOUS NURSING SCHEDULES

schedules		factor values (x 10^{-1})				
name	number	C	O	P	H	T
a	6	6.0	5.5	6.2	5.9	6.6
b	8	5.6	4.2	5.5	9.9	6.0
c	3	9.9	6.3	5.8	4.8	6.6
d	12	4.4	3.7	9.9	5.9	6.0
e	5	6.0	6.0	4.5	4.8	6.6
f	14	4.3	9.9	5.5	5.9	3.7
g	1	6.0	4.2	6.2	4.9	9.9
h	10	9.9	4.6	4.5	5.9	4.5
i	7	5.6	4.1	4.8	9.9	4.7
j	2	4.9	4.2	9.9	5.0	6.6
k	11	4.8	9.9	4.8	5.9	4.5
l	15	9.9	4.2	6.2	4.8	4.0
m	4	4.0	6.3	4.6	4.9	9.9
n	13	4.0	6.3	4.6	4.9	4.1
o	9	4.4	3.7	4.8	5.9	3.1

schedulers will differ in the summation weights per quality factor. An unbalanced set could therefore affect the determination of the summation weights.

Furthermore, the thirty-six unacceptable factor values were distributed ‘normally’ over the fifteen fictitious nursing schedules. As there are five quality factors, a nursing schedule has five or less unacceptable factor values. This number of unacceptable factor values per nursing schedule is abbreviated to ‘UFV/sched’ in table 6.1.

In total, there are thirty-two different nursing schedules possible considering only the acceptable and unacceptable values for the five quality factors. The theoretically possible number of different nursing schedules per number of unacceptable factor values is abbreviated to 'Psched/UFV' in table 6.1.

Fifteen of these thirty-two possible combinations of acceptable and unacceptable factor values were used as a design for arranging the fictitious nursing schedules. The number of arranged nursing schedules per number of unacceptable factor values is abbreviated to 'Asched/UFV' in table 6.1.

Table 6.1 shows the 'normal' distribution of the thirty-six unacceptable factor values over the fifteen fictitious nursing schedules. This table shows that this distribution for the arranged nursing schedules (Asched/UFV) is similar to the normal distribution for the theoretically possible number of different nursing schedules per number of unacceptable factor values (Psched/UFV).

The fifteen nursing schedules were arranged according to both distributions. Table 6.2 shows the factor values for each of these fifteen nursing schedules.

Table 6.3 CORRELATION COEFFICIENTS FOR THE FACTOR VALUES

	O	P	H	T
C	- 0.22	- 0.11	- 0.13	- 0.08
O		- 0.35	- 0.21	- 0.16
P			- 0.14	0.17
H				- 0.22

The requirement of a balanced set of nursing schedules can also be tested on the basis of the correlation coefficients between the values of the five quality factors. Table 6.3 shows these correlation coefficients for the arranged fifteen nursing schedules.

Table 6.4 shows the probability that a pair of quality factors for these fifteen nursing schedules do not correlate. As all probabilities are above an error tolerance of five percent (0.05), the values of the five quality factors are not significantly correlated.

Summarizing, the set of fifteen fictitious nursing schedules is balanced with respect to the distribution of the unacceptable factor values over both the quality factors and the nursing schedules. Furthermore, no pair of quality factors is correlated.

Table 6.4 PROBABILITIES OF NON-CORRELATION

	O	P	H	T
C	0.44	0.69	0.63	0.77
O		0.20	0.45	0.58
P			0.61	0.54
H				0.43

6.2 RESULTS OF THE AUDITING EXPERIMENT

Five nurse schedulers were asked to give a total quality mark ($Q_{s,i}$), on a scale from one to ten, to each of the fifteen fictitious nursing schedules of the test set discussed above. The nurse schedulers were instructed to a give an unacceptable nursing schedule a mark below 5.5. Furthermore, the nurse schedulers were asked to clock the amount of time needed for the audit of each of these fifteen nursing schedules. Table 6.5 shows the quality marks given to the fifteen nursing schedules by the five nurse schedulers (i.e. the results of the auditing experiment).

6.2.1 Determination of summation weights

The given quality marks were used to validate the hypothesis of robustness. To test this hypothesis, the results of the auditing experiment were reformulated as an overdetermined set of linear equations in the form of $Ax \approx b$, where A represents the $m \times n$ matrix ($m \geq n$) of factor values (table 6.2) and b represents the quality marks given to each of the fifteen nursing schedules by one of the five nurse schedulers (each column from table 6.5). The values per nurse scheduler of the column x, which minimizes the Euclidean norm of the residual vector $r=b-Ax$, are determined by using a NAG FORTRAN Library Routine, called F04JAE, for finding the minimal solution to the overdetermined linear least-squares problem $Ax \approx b$ (Lawson & Hanson, 1974, pp. 180-198). Table 6.6 gives the resulting least-squares solution per nurse scheduler. For each of these solutions, a rank of five was computed.

Table 6.5 QUALITY MARKS GIVEN BY THE NURSE SCHEDULERS ($Q_{s,r}$)

nursing sched-ules	nurse schedulers				
	1	2	3	4	5
a	7.5	7.5	9.8	7.5	6.2
b	5.4	5.5	8.8	7.5	4.5
c	6.5	7.0	9.4	6.5	6.0
d	4.5	1.0	7.8	6.0	4.3
e	5.4	5.0	8.6	6.0	5.0
f	3.5	2.0	8.0	6.5	5.8
g	5.4	7.0	9.2	7.8	7.5
h	5.0	3.0	7.8	6.5	4.3
i	4.0	4.5	8.8	6.2	4.5
j	5.4	5.0	9.0	7.0	4.5
k	6.0	3.0	7.0	5.5	5.5
l	5.4	5.0	8.4	6.0	3.8
m	3.0	1.0	8.2	5.4	4.5
n	3.5	4.0	6.6	5.4	5.2
o	4.0	5.0	6.8	5.4	4.0

The resulting least-squares solutions show that nurse schedulers strongly differ concerning their personal summation weights. Nurse scheduler number two put most of the weight on completeness, while two other nurse schedulers (i.e. numbers three and four) distributed these weights more equally over all five quality factors. And nurse scheduler number five considered optimality and continuity to be the most important quality factors, while for nurse scheduler number one, the most important quality factors turned out to be completeness and proportionality.

Furthermore, these results show one negative weight value. An investigation of this negative summation weight of the second scheduler for the optimality factor showed that this nurse scheduler ‘rewarded’ qualitative overstaffing. This was probably due to the fact that this nurse scheduler's own nursing unit only consisted of nurses with the same level of nursing expertise. This nurse scheduler was therefore unfamiliar with differences in optimality. As this factor was unknown to this second nurse scheduler, the results of this nurse scheduler were analyzed with the requirement that all solution components must be non-negative. This was done by using another NAG FORTRAN Library Routine, called E04NCE. Table 6.6 also shows the resulting non-negative least-squares solution for this second nurse scheduler (2^+). This solution also shows the weight nurse scheduler number two puts on the completeness factor.

Table 6.6 SUMMATION WEIGHTS PER NURSE SCHEDULER

scheduler	summation weights					
	ω^c	ω^o	ω^p	ω^h	ω^t	σ
1	3.3	1.3	2.1	0.8	1.0	1.1
2	4.2	- 1.1	0.0	1.9	2.3	2.2
3	3.0	1.6	2.6	3.3	3.6	0.7
4	2.0	1.2	2.4	3.0	2.3	0.7
5	1.0	3.1	0.8	0.8	3.0	0.9
2^+	3.9	0.0	0.0	2.0	1.5	

6.2.2 Goodness of fit

Table 6.6 also shows the standard error (σ) for each least-squares solution. This standard error has the value $\sqrt{(r^T r/(m-k))}$ when $m>k$, and the value zero when $m=k$. The $r^T r$ in this computation expresses the residual sum of squares. Furthermore, m represents the number of rows of the matrix A, which is 15 in this case, and k stands for the rank of matrix A, which turned out to be 5 for all solutions. Therefore, in this case, the standard error σ has the value of $\sqrt{(r^T r/10)}$.

These standard errors provide a measure for the quality of the least-squares solutions. With a probability of ninety-five percent, the quality mark a nurse scheduler will give (or has given) to a nursing schedule ($Q_{s,i}$) will be above $QN_{s,i}$-2σ

Table 6.7 PREDICTED QUALITY MARKS ($Q'_{s,i}$)

nursing schedules	nurse schedulers					
	1	2	3	4	5	2^+
a	5.1$^-$	4.6$^-$	8.6	6.6	5.3	4.5$^-$
b	4.9	5.2	9.2	7.3	4.9	4.9
c	6.3	5.9	9.4	7.1	5.8	5.9$^-$
d	5.1	3.9$^-$	8.6	6.9	4.7	3.8$^-$
e	4.7	4.3	7.9	6.0	5.2	4.4
f	4.7$^-$	2.7	7.6	6.0	5.5	3.3$^-$
g	5.2	5.3$^-$	9.3	6.9	5.8$^-$	5.1$^-$
h	5.7	5.8$^-$	8.4	6.4	4.6	5.6$^-$
i	4.7	4.9	8.5	6.8	4.4	4.6
j	5.3	4.1	8.7	6.9	5.0	4.0
k	4.8$^-$	3.1	7.8	6.1	5.8	3.7
l	5.9	5.5	8.3	6.3	4.3	5.4
m	4.5$^-$	4.2$^-$	8.6	6.4	6.1$^-$	4.3$^-$
n	3.9	2.9$^-$	6.5	5.1	4.3	3.1
o	3.7	3.3$^-$	6.2	5.0	3.4	3.2$^-$
out of range	4	7	0	0	2	8
total out of range	13					14

$^-$ $p > 0.05$

and below $Q'_{s,i}+2\sigma$. This $Q'_{s,i}$ is the predicted quality mark on the basis of the summation of the factor values multiplied by the nurse scheduler's personal weight values. For example, on the basis of the weight values of the fifth nurse scheduler, a quality mark of 6.1 ± 1.0 (i.e. two times the tolerable standard error of 0.5) is predicted for schedule m with a probability of ninety-five percent. The actual quality mark given to this schedule by this nurse scheduler was 4.5, which is not within the predicted range.

For all the fictitious nursing schedules, table 6.7 shows the predicted quality marks for each of the five nurse schedulers. When applying a tolerable standard error of 0.5 for all schedulers, which has been assumed to be realistic for a marking scale from one to ten, not all the predicted quality marks fall within the ninety-five percent probability range. Those that do not fall within this range are marked with a negative sign ($^-$).

Table 6.8 MEAN MARKING SCORES AND THE STANDARD DEVIATION PER NURSE SCHEDULER

	nurse schedulers					mean
	1	2	3	4	5	
$Q^m_{s,i}$	5.0	4.4	8.3	6.4	5.0	5.8
sd_i	1.2	2.0	0.9	0.8	1.0	1.2

These predicted quality marks illustrate the values of the standard errors shown in table 6.10. The hypothetical modelling of nursing schedule quality as a weighted sum of the factor values is strongly supported by the quality marks given by two of the five nurse schedulers (nurse schedulers numbers three and four). All quality marks given by these two nurse schedulers are explained by the hypothetical model as represented in formula 6.1. The quality marks given by another two of the five nurse schedulers (nurse schedulers numbers one and five) still support the presented model, although not very strongly. However, the model is not supported by one of the five nurse schedulers (nurse scheduler number two). In total, sixty-two of the seventy-five quality marks given by the five nurse schedulers to the fifteen nursing schedules support the hypothetical model. This means that eighty-three percent of these quality marks can be explained on the basis of a weighted sum of the factor values. The restriction that all summation weights must be non-negative resulted in a total of fourteen out-of-range predictions, which slightly reduces this percentage to eighty-one.

Table 6.9 LINEARLY TRANSFORMED QUALITY MARKS ($Q^{n}_{s,i}$)

nursing schedules	nurse schedulers				
	1	2	3	4	5
a	7.6	7.0	7.1	6.9	6.7
b	5.9	6.1	6.1	6.9	5.0
c	6.7	6.8	6.7	5.7	6.5
d	5.3	3.9	5.0	5.1	4.8
e	5.9	5.8	5.8	5.1	5.5
f	4.4	4.3	5.2	5.7	6.3
g	5.9	6.8	6.5	7.3	8.0
h	5.5	4.8	5.0	5.7	4.8
i	4.7	5.6	6.0	5.3	5.0
j	5.9	5.8	6.3	6.3	5.0
k	6.3	4.8	4.1	4.4	6.0
l	5.9	5.8	5.6	5.1	4.3
m	3.9	3.9	5.4	4.3	5.0
n	4.3	5.3	3.9	4.3	5.7
o	4.7	5.8	3.9	4.3	4.5

6.2.3 Coping with differences in marking style

Nurse schedulers might differ with respect to their marking style. This means that both the mean marking score ($Q_{s,i}^{m}$) and the standard deviation (sd_i) can differ per nurse scheduler i. The five nurse schedulers who participated in the ranking experiment did strongly differ with respect to both the mean marking score and the standard deviation of these scores. Table 6.8 shows these differences.

By applying a standard transformation, these differences in marking style can be eliminated. This standard transformation results in a standard distribution for each nurse scheduler. A standard distribution has a mean of zero and a standard deviation of one. The computation of the original marking values ($Q_{s,i}$) into standard marking values ($Q_{s,i}^s$) is based on the equation of $Q_{s,i}=Q_{s,i}^m+sd_i \times Q_{s,i}^s$. In order to achieve new marking values ($Q_{s,i}^n$) with a mean of 5.5, which is the mean of the range from one to ten, and a standard deviation of one, the following linear transformation was performed.

$$Q_i^n = \frac{Q_i - Q_i^m}{sd_i} + 5.5$$

Formula 6.2: Linear transformation of the original marking values into new marking values

Table 6.10 NEW SUMMATION WEIGHTS PER NURSE SCHEDULER

scheduler	summation weights					
	ω^c	ω^o	ω^p	ω^h	ω^t	σ
1	3.1	1.6	2.3	1.3	1.2	0.9
2	3.0	0.7	1.2	2.5	2.0	1.2
3	2.3	0.4	1.6	2.1	3.1	0.7
4	1.8	0.5	2.1	2.7	2.3	0.8
5	1.2	3.3	0.9	1.0	3.1	0.9

By applying this formula, the original marking values ($Q_{s,i}$) were linearly transformed into the new marking values ($Q_{s,i}^n$). Table 6.9 shows these linearly transformed quality marks.

Table 6.10 gives the resulting least-squares solution per nurse scheduler. Also for each of these solutions, a rank of 5 was computed (i.e. A is of full rank).

For all the fictitious nursing schedules, table 6.11 shows the predicted new quality marks (QN_i^n) for each of the five nurse schedulers. When applying a tolerable standard error of 0.5 for all schedulers, not all the predicted quality marks fall within the ninety-five percent probability range. Those that do not fall within this range are marked with a negative sign ($^-$).

Table 6.11 shows that in fourteen of the seventy-five cases the predicted new quality marks are out of range. In thirteen of the seventy-five cases, the predicted original quality marks are out of range. Although the linear transformation did have a small effect on the goodness of fit per nurse scheduler, these results show that differences in marking style do not affect the overall results of the auditing experiment. These overall results show that more than eighty percent of the quality marks given to fifteen fictitious nursing schedules by five nurse schedulers fall within the predicted range.

6.3 CONCLUSIONS OF THE AUDITING EXPERIMENT

The results of the auditing experiment show that more than eighty percent of the quality marks given to fifteen fictitious nursing schedules by five nurse schedulers can be explained on the basis of a weighted sum of factor values. These results support the hypothesis of robustness. Furthermore, these results are not affected by differences in marking style. Therefore, the auditing experiment shows that nursing schedule quality can be modelled as a weighted sum of factor values, where the factor definitions are generic and the summation weights are specific. This provides a positive answer to the third research question asked in the third chapter (i.e. "Can the total nursing schedule quality be explained on the basis of a weighted sum of factor values?"). The next chapter describes the investigation of the application of this operational model of nursing schedule quality.

Table 6.11 PREDICTED LINEARLY TRANSFORMED QUALITY MARKS ($Q'^{n}_{s,i}$)

nursing sched-ules	nurse schedulers				
	1	2	3	4	5
a	5.7 -	5.7 -	5.9 -	5.8 -	5.7
b	5.7	6.3	6.3	6.4	5.4
c	6.8	6.6	6.5	6.1	6.3
d	5.7	5.4 -	5.8	6.0	5.1
e	5.3	5.3	5.4	5.1	5.6
f	5.4	4.9	4.7	4.9	6.0
g	5.9	6.0	6.6	6.2 -	6.2 -
h	6.1	6.2 -	5.8	5.6	5.1
i	5.3	6.0	5.8	6.0	4.9
j	5.9	5.5	6.0	6.0	5.4
k	5.5	5.1	4.9	5.0	6.3
l	5.3	6.0	5.7	5.5	4.9
m	5.1 -	5.4 -	6.0	5.6 -	6.5 -
n	4.4	4.2 -	4.2	4.3	4.7
o	4.2	4.3 -	4.1	4.3	3.7
out of range	2	6	1	3	2
total out of range	14				

- $p > 0.05$