British Journal of Anaesthesia 1992; **68**: 32–38

EFFECTS OF POSTURE, PHONATION AND OBSERVER ON MALLAMPATI CLASSIFICATION

E. J. THAM, C. D. GILDERSLEVE, L. D. SANDERS, W. W. MAPLESON AND R. S. VAUGHAN

SUMMARY

We have studied the effects of phonation and posture on the Mallampati classification of view of the pharyngeal structures. Differences between observers were allowed for by the experimental design and log–linear modelling. Sixty-four patients were assessed on the ward, sitting upright, with and without phonation, by each of two observers. Another 64 patients were assessed without phonation, but both upright and supine, again by both observers. Phonation (the patient saying "Ah") produced a marked, systematic improvement of view; moving to the supine posture produced a small, systematic, non-significant worsening of the view. Differences between observers were non-systematic but substantial. About 25% of patients phonated spontaneously. It is recommended that anaesthetists make their own assessments of Mallampati classification, with the patient in either of the postures but always either with or without phonation, and thereby gradually "calibrate" their assessments against the degree of difficulty encountered in intubation.

KEY WORDS

Intubation, tracheal: prediction of difficulty.

Unexpected difficulty with tracheal intubation is a significant contributor to morbidity and mortality in clinical practice [1]. Mallampati and his colleagues [2, 3] demonstrated that a possible difficult tracheal intubation could be predicted before operation using a simple grading system involving the ability to visualize pharyngeal structures. To the three classes described by Mallampati, a fourth was added by Samsoon and Young [4]. Both the Mallampati and Samsoon and Young assessments were performed with the patient in the sitting position.

During clinical practice, situations may arise in which it is not feasible for the patient to sit up for assessment of the airway. Initially, the aim of this study was to determine if the Mallampati class was different with the patient in the sitting and supine postures. However, it was noted at an early stage that several patients were phonating spontaneously (saying "Ah") during the inspection, and that this appeared to change the view of the pharyngeal structures. Therefore a study was undertaken to determine the effects of phonation on the Mallampati classification before proceeding to the original study on the effect of posture.

PATIENTS AND METHODS

We studied patients older than 16 yr, admitted for elective surgical procedures. All gave their informed consent. Patients who were in their immediate postoperative period were excluded.

Assessment

The airway was assessed according to the pharyngeal structures seen, using the method described by Mallampati [3] with the modification of Samsoon and Young [4]:

Class 1—soft palate, fauces, uvula and pillars visible.
Class 2—soft palate, fauces, uvula visible.
Class 3—soft palate, base of uvula visible.
Class 4—none of the soft palate visible.

In the phonation study, the patient sat upright, with the head in the neutral position, the mouth opened maximally and the tongue protruded maximally. The observer was seated opposite the patient at eye level and the pharyngeal structures were viewed with a torch. One assessment of the class of view was made with the patient not phonating and another with the patient phonating.

In the posture study, one assessment was made as in the phonation study (but without phonation); the other was made with the patient supine with the head on one pillow, and with the observer looking vertically downwards.

Study design

Sixty-four patients were included in each of the two main studies. Although the main object was determination of the effects of phonation and posture, it was thought that there might be differences between observers. Therefore all assessments were made by both of the same two observers (C.D.G. and E.J.T.) throughout all the studies.

E. J. THAM, M.B., B.CH., F.C.ANAES.; C. D. GILDERSLEVE*, M.B., B.CH., F.C.ANAES.; L. D. SANDERS, PH.D., A.F.B.PS.S., C.PSYCHOL.; W. W. MAPLESON, D.SC., F.INST.P.; R. S. VAUGHAN, M.B., B.S., F.C.ANAES.; Department of Anaesthetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN. Accepted for Publication: July 11, 1991.

* Present address: Princess of Wales Hospital, Bridgend, Mid Glamorgan CF31 1RQ.

TABLE I. *Assignment of observers and conditions to groups of patients in each study. C = Control condition (upright, not phonating); E = experimental condition (phonating or supine); Obs = observer. Note that, in the serial numbers of sets of (eight) patients, 1–4 refer to different wards; a and b refer to the physically separate halves of each ward*

Serial number of set		Assessment on each patient			
Study 1	Study 2	1	2	3	4
1a, 3a	1a, 4a	Obs1 C	Obs2 E	Obs1 E	Obs2 C
1b, 3b	1b, 4b	Obs2 C	Obs1 E	Obs2 E	Obs1 C
2a, 4a	2a, 3a	Obs1 E	Obs2 C	Obs1 C	Obs2 E
2b, 4b	2b, 3b	Obs2 E	Obs1 C	Obs2 C	Obs1 E

Thus, in the main studies, each patient was assessed four times, by each of the two observers (independently) under each of the two conditions (phonation/no phonation or upright/supine).

The four assessments were made in four different orders according to a latin square design [5] (table I). Thus, at the start of each main study, observer 1 made the first assessment of the eight patients in set 1a (in one half of ward 1), under "control" conditions (no phonation, upright) ("Obs1 C" in table I). At the same time, observer 2 assessed the eight patients in set 1b in the opposite (physically separate) half of ward 1 (Obs2 C), also under control conditions. Then the observers exchanged locations and assessed the eight patients in the half of the ward which they had not yet visited, under "experimental" conditions (phonation or supine)—and so on, as shown in table I. The observers did not refer to each other's assessments, or to their own first assessment when making the second assessment on each patient.

This experimental design achieved the following objectives. Each patient acted as his or her own control for differences of both condition and observer. All patients were observed under control conditions before experimental conditions, and under experimental conditions before control conditions. Fifty percent of the patients were seen by observer 1 first, and 50% by observer 2 first. Each observer assessed 16 patients once before returning for the second assessment; this minimized any memory of the class assigned at the first assessment.

Subsidiary study

Another 100 patients were examined, 50 by each observer, to determine what proportion of patients phonated spontaneously on being asked to open the mouth and protrude the tongue. The observer noted if the patient phonated, mimed phonation (flattened the tongue and contracted the paired levator veli palatine muscles), or did not phonate.

Statistical analysis

The results were analysed by log–linear modelling [6] using the statistical package GLIM version 3.77 (plus MINITAB release 7.2 for exact P values) running on a Digital VAX mainframe computer under the VMS operating system, version 5.3-1. Log–linear modelling may be regarded as an extension of multiple linear regression to ordinal categorical data. It has been used here to estimate the nature and magnitude of the effect of the explanatory factors, condition and observer, and their interaction, on the distribution of the assessments between the different Mallampati classes.

Although log–linear modelling is now a standard technique, it is relatively new and the explanations in text books are not easy to follow. Therefore, for those readers who may wish to use it, or simply to satisfy themselves as to how it works, an explanation of how it has been used to produce the results presented in the body of the paper is given in Appendix 1. Appendix 2 describes some additional features of log–linear modelling which were applied also to the present data but did not reveal any finding of clinical importance.

Readers who are interested only in the clinical implications of the results will find all the essential information in the main text.

RESULTS

The observers reported that, when they returned to a patient for the second assessment of the airway, under the alternative conditions, they could not remember the class of view assigned at the first assessment.

Phonation had a marked effect on the distribution of assessments between Mallampati classes (table II): the assessments were "shifted" systematically towards the better classes (1 and 2) relative to no phonation. In contrast, posture had only a very small effect: changing from the upright to the supine position produced a small shift in the opposite direction.

The "expected" frequencies (numbers in parentheses in table II) are those predicted by the log–linear model on the assumption that there is *only* a systematic effect. The closeness of the fit to the observed frequencies in the phonation study indicates that most of the effect of phonation was indeed systematic. The term "fitted odds ratio" is explained

TABLE II. *Two-way contingency tables showing the effects of condition on the observed and (fitted) distributions of assessments between the four Mallampati classes, with the fitted odds ratios (ratio of odds of being in the upper of two adjacent classes under the experimental condition (phonation or upright) to the odds under the control condition, for example (49.0/59.3)/(20.0/57.7) = 2.39)*

Condition	Mallampati class 1	2	3	4	Totals	Fitted odds ratio
Phonation study						
Phonation	49 (49.0)	60 (59.3)	17 (18.3)	2 (1.4)	128	2.39
No phonation	20 (20.0)	57 (57.7)	44 (42.7)	7 (7.6)	128	
Totals	69	117	61	9	256	
Posture study						
Supine	29 (29.1)	48 (46.3)	41 (44.1)	10 (8.5)	128	0.86
Upright	35 (34.9)	46 (47.7)	42 (38.9)	5 (6.5)	128	
Totals	64	94	83	15	256	

TABLE III. *Two-way contingency table showing the effects of observer on the observed and (fitted) distributions of assessments between the four Mallampati classes, with the fitted odds ratios*

Observer	Mallampati class 1	2	3	4	Totals	Fitted odds ratio
Phonation study						
1	34 (29 3)	49 (58.2)	39 (34.8)	6 (5 8)	128	
						0.75
2	35 (39.7)	68 (58.8)	22 (26.2)	3 (3.2)	128	
Totals	69	117	61	9	256	
Posture study						
1	38 (32.6)	40 (47.1)	39 (41.0)	11 (7.3)	128	
						1.03
2	26 (31.4)	54 (46.9)	44 (42.0)	4 (7 7)	128	
Totals	64	94	83	15	256	

TABLE IV. *Number of instances of each type of agreement and disagreement between observers in their assignment of patients to the four Mallampati classes. Numbers for agreement are italicized: a total of 73 (57% of all assessments by each observer) in the phonation study; 76 (59%) in the posture study*

Observer 2 Mallampati class	Observer 1 Mallampati class 1	2	3	4	Totals
Phonation study					
1	*20*	12	3	—	35
2	13	*35*	18	2	68
3	—	2	*17*	3	22
4	1	—	1	*1*	3
Totals	34	49	39	6	128
Posture study					
1	*22*	4	—	—	26
2	13	*28*	13	—	54
3	3	8	*24*	9	44
4	—	—	2	*2*	4
Totals	38	40	39	11	128

in the heading to table II and, more fully, in Appendix 1, but it may be noted here that the difference of the ratio from unity indicates the magnitude of the systematic effect: large and significant for phonation (ratio = 2.39 (table II); $P < 0.0001$, see Appendix 1) and small and non-significant for posture (ratio = 0.86; $P = 0.3$).

The differences between the two observers showed no obvious systematic effect in either study (table III): the fitted odds ratios are close to unity and non-significant (0.75, $P = 0.06$, in the phonation study; 1.03, $P = 0.8$, in the posture study). In contrast, the large differences between the observed and "expected" frequencies in each table indicate that there was a large amount of non-systematic variation. Even this does not reveal all the disagreement between observers. For instance, in the phonation study, the 39 class 3 assessments by observer 1 do not include all the 22 class 3 assessments by observer 2. In fact, in only 17 instances did the observers agree on a class 3 assessment. Table IV shows all the details of agreement and disagreement between observers in each study. Agreement occurred in only just over 50% of the instances: 73 or 76 of a total of 128 assessments by each observer. The disagreements were mostly over a difference of only one class, but there was one instance in the phonation study of disagreement between classes 1 and 4!

"Interaction" between observer and condition (the effect of condition being greater in one observer than the other) was negligible in the posture study and did not reach significance in the phonation study (see Appendix 1). Additional analysis (Appendix 2) failed to reveal any "learning" effect in the observers.

In the subsidiary study, there was no evidence of phonation in 77 of the 100 patients, 22 mimed phonation and one said "Ah" aloud.

DISCUSSION

The present study was concerned with revealing how the assignment of patients to the Mallampati classes (and hence presumably the predictive power of the test) might be modified, primarily by change of posture, but also by phonation and by differences between observers.

The small, non-significant effect of posture in this sample of patients suggests that assessment of patients placed in the supine position may be interpreted in the same way as assessment of those in the upright position. However, it is shown in Appendix 1 that the results for the present sample of patients are consistent (lower 95% confidence limit of the systematic effect of posture) with an appreciable shift of assessments in the population at large towards the worst class in the supine posture compared with the upright posture. This "worst probable case" is illustrated in table V.

It seems plausible that the effect of phonation, first reported by Wilson and John [7], should be almost entirely systematic, as we have found: the action of saying "Ah" may be expected to produce a consistent improvement in the view of the pharyngeal structures. Furthermore, a study of 334 patients by Oates and others in 1990 [8] showed a similar improvement of view with phonation: when we fitted a log-linear model to their data, the resulting odds ratio was lower than, but consistent with, ours: 2.02 with 95% confidence limits of 1.58–2.58 compared with our 2.39. As in our study, there was little residual non-systematic variation associated with phonation.

As nearly 25% of the patients in the subsidiary study attempted phonation, thereby improving the

TABLE V. *Two-way contingency table showing the most extreme likely effect of posture on Mallampati class, calculated by combining the lower 95% confidence limit of the odds ratio (0.649) with the observed total frequency in each class*

Condition	Mallampati class 1	2	3	4	Totals	Chosen odds ratio
Supine	23.9	45.1	48 7	10.3	128	
						0.649
Upright	40 1	48.9	34 3	4.7	128	
Totals	64	94	83	15	256	

view of the pharyngeal structures, it seems that the results of any study which fails to control phonation must be interpreted with caution.

The differences between observers in the present study were mostly non-systematic and did not improve from the phonation to the posture study (table IV).

A striking difference between the results of our study and those of others is the different distribution of patients between the Mallampati classes. In the original study by Mallampati and colleagues [3], where "class 3" would have included the later Samsoon and Young [4] class 4, the assignments were roughly 70%, 20% and 10%, to classes 1, 2 and 3, respectively, while in both the 1990 and the 1991 studies of Oates and others [8, 9], they were approximately 70%, 10% and 20%. In contrast, in our study, in the absence of phonation and combining classes 3 and 4, the assignments were 20%, 40% and 40%. Another difference is that the disagreement between observers in our study amounted to 42% of all assessments (table IV) whereas, in the study of Wilson and John [7] it was 34%, and in the 1991 study of Oates and others [9] it was only 15% overall. A third difference was that, whereas our disagreements were very largely non-systematic, Oates and others [9] found that one of their observers was systematically different from the other three, assigning 45%, 10% and 45% of patients to classes 1, 2 and 3, respectively, against 80%, 10% and 10% for the total of the other three observers.

One possible interpretation of these differences is that, although our two observers differed from each other in their assessments in a largely non-systematic way, they may both have given systematically worse assessments than other observers. The explanation given by Oates and others [9] for their systematic inter-observer difference was that their one anomalous observer was failing to persuade his patients to open their mouths and protrude their tongues maximally. To test if this might be the explanation for our results we asked our two observers that, after performing each routine Mallampati examination in their clinical work, they should urge the patient to try harder and note any change in Mallampati class. In 25 consecutive patients each, they found that only a few patients could open their mouths further, and in none was the view improved. However, the distribution of these 50 patients between the Mallampati classes (40%, 50% and 10% for classes 1, 2 and 3) was shifted part of the way from the pattern in our main study (20%, 40% and 40%) to that in the studies of Mallampati and Oates and their colleagues [3, 8, 9] (70%, 15% and 15%). This suggests that, during our main study, the observers may not have been persuading the patients to co-operate fully, even though they both believed that they were being more thorough than in their previous clinical practice. If so, this would suggest that, in teaching the Mallampati test to trainee anaesthetists, great stress should be laid on explaining the nature of the test to patients in order to obtain their full co-operation.

In the last 50 patients, the distribution (40%, 50% and 10%) was still substantially different from that in the patients from Glasgow [8, 9] or Boston [3] (70%, 15% and 15%). Maybe there are genuine geographic differences in distribution.

A wider issue, not addressed directly by the present study, is the power of the Mallampati test to predict difficult or failed intubations. Such intubations make a significant contribution to mortality and morbidity: the *Confidential Enquiry into Maternal Deaths 1982–84* [10] reported 10 deaths during obstetric anaesthesia associated with difficulty in intubation. Therefore it is important to be able to predict such difficulties.

A large tongue is known to be a cause of difficulty; therefore it is plausible to look inside the mouth to see how much it obscures the view of the pharynx—in the manner codified by Mallampati and colleagues [3]. However, the technique has proved disappointing in practice. Previous studies, using only the three original Mallampati classes, have found that the occurrence of "class 3" predicted only 14 of 28 cases of "inadequate exposure of the glottis" [3] and one of six [8] or five of 12 [9] "difficult" laryngoscopies.

This raises the question, "Can the predictive power of the Mallampati test be improved?". The present study shows that the supine position probably does not worsen the predictive power. Also, our study, combined with the studies of Wilson and John [7] and Oates and others [8], shows that, for consistent classification, it is essential to standardize on either phonation or non-phonation. Phonation gives a notably better view of the pharynx on average. Therefore the view with phonation might possibly be a better predictor of difficulty in intubation. This would be true if the patients in whom the view improved from class 3 or 4 to class 1 or 2, were those who presented little or no difficulty in intubation. In the 1990 study of Oates and others [8], the view of the pharynx was improved in 43 of 66 patients in class 3 (including class 4), and 42 of these did not present difficulty in intubation. In contrast, the remaining patient of the 43 showing improvement was the only "class 3" who did present difficulty. Thus phonation removed 67% of the false positives—but also removed one true positive. However, as the authors noted, a very large study would be needed to provide a definitive answer to the question, because of the rarity of difficulty in intubation.

Perhaps the greatest obstacle to improving the predictive power of the test is the great variability of the results within and between studies. For instance, if there is a genuinely different distribution between Mallampati classes in Cardiff from that in Glasgow or Boston, is there a matching difference in the incidence of difficult intubations? If not, a different "conversion" would be necessary from Mallampati class to "difficult intubation". Second, if failure to elicit full co-operation of the patient is a major source of variation, does random variation in the degree of effort by the patient explain most of the non-systematic difference between observers—or do observers disagree over the classification of a particular view of the pharynx?

TABLE VI. *Analyses of deviance for the two studies*

	Phonation study					Posture study			
	Residual		Change in				Change in		
Terms in model	df	Deviance	df	Deviance	P	Residual deviance	df	Deviance	P
Minimal model	10	44.16				11.08			
Effect of condition (phonation or posture)									
Systematic	9	16.65	1	27.51	< 0 0001	9.94	1	1.14	0 29
Non-systematic	7	16.19	2	0.46	0.79	8.76	2	1.18	0.55
Totals			3	27.97	< 0.0001		3	2.32	0 51
Difference between observers									
Systematic	6	12.69	1	3 50	0.061	8 71	1	0 05	0.82
Non-systematic	4	7.26	2	5.43	0.066	0.70	2	8.01	0.018
Totals			3	8.93	0.030		3	8.06	0.045
Interaction between condition and observer									
Systematic	3	5.33	1	1.93	0.16	0.18	1	0.52	0.47
Non-systematic	0	0.00	2	5 33	0.07	0 00	2	0.18	0.91
Totals			3	7.26	0.064		3	0.70	0.87

It is difficult to see what the individual observer may do to improve his prediction of difficult intubation. As our two observers showed just as much disagreement in the second (posture) study as in the first (table IV), mere practice is unlikely to reduce random variation.

If one observer's classification is systematically different from others (as in the 1991 study by Oates and others [9]) he could perhaps correct this, either by eliciting fuller co-operation from the patient (if that is the explanation) or else by adjusting his interpretation of what he sees; but how does he discover his bias in the first place, unless he participates in a comparative trial with other anaesthetists? In the long term, he might "calibrate" his classification against his experience of difficulty in intubation but, again because of the rarity of such difficulties, that may take some years and, if he is inherently variable in his classifications, for whatever reason, his predictions may probably never be very reliable. Certainly, a Mallampati class assigned by one anaesthetist is of limited value to another unless the first anaesthetist's "calibration" is known—and known to be reliable.

Perhaps the next investigation in this field should be designed to distinguish between the different causes of variation in classification, and in the relationship between classification and difficulty in intubation. This might be achieved by photographing the patient's pharynx, first after the standard instruction to "open the mouth as wide as possible and protrude the tongue as far as possible", and again after exhorting the patient to try harder; and for more than one observer to do this with each patient, preferably in some suitable cross-over design involving several observers, and in patients in different parts of the country and of the world. Then the photographs could be assigned to the different classes by each of several observers, at first serially—looking at one photograph after the other—and then by spreading out all the photographs and arranging them in rank order before assigning classes. If, in addition, all these patients were followed through to laryngoscopy, it might be possible to distinguish some common feature in the photographs of those patients in whom intubation of the trachea was difficult or impossible—perhaps in a way not included in the present Mallampati classes. If so, it would be logical to reduce the grading of the view of the pharynx to just two classes: likely and not likely to be difficult to intubate; also, the set of photographs would provide a valuable training resource. If not, the charge might be made that a great deal of resources had been wasted.

In the meantime, we may conclude that performing the Mallampati test with the patient supine is unlikely to make much difference to the class assigned but that, to be consistent, the anaesthetist should ensure that the patient does not phonate.

APPENDIX 1

EXPLANATION OF THOSE BASIC FEATURES OF LOG–LINEAR MODELLING USED IN THE MAIN TEXT

A log–linear model consists of an equation which predicts the *logarithms* of the "expected" frequencies in the cells of a contingency table and is a *linear* combination of terms (which may themselves be non-linear functions of explanatory factors)—hence *log–linear* modelling. It quantifies the effects of the explanatory factors on the outcome variable (Mallampati class in the present case) in two ways. One way is in terms of the fitted odds ratios mentioned in the Results, each ratio corresponding to a term in the equation. The other way is in terms of the "deviance" (see below) of the observed frequencies in the contingency table from the "expected" frequencies—that is, the frequencies fitted, or predicted, by the current model. Therefore, log–linear modelling is related to the familiar chi-square test for a contingency table, but it can take account of all the explanatory factors at once and include interactions between them. Therefore, it is also an extension of multiple linear regression to categorical data.

In a "minimal model", the only terms in the equation are those necessary and sufficient to make the marginal totals of the table of predicted frequencies agree with the marginal totals for the observed frequencies. For instance, in table II, as there are equal totals for each condition, the totals for each Mallampati class would be shared equally between the two conditions. Thus, for both conditions in the phonation study, the "expected" frequencies for classes 1–4 in a minimal model would be 34.5, 58.5, 30.5 and 4.5, respectively. This is necessarily accompanied by an odds ratio of 1 between all adjacent pairs of classes.

TABLE VII. *Two-way contingency table for the phonation study, showing how the distribution of assessments between Mallampati classes changes with the serial number of each set of 16 assessments by the observers (and with ward number). The fitted odds ratios and expected frequencies were obtained from a model with two terms for the systematic effect of serial number: one linear and one quadratic in serial number. Note that, for instance, the first set of 16 assessments by each observer corresponds, not to the first row of table I, but to the first two assessments on the 16 patients in ward 1 (the 1a and 1b sets of eight patients in table I)*

Ward no.	Serial number of set	Mallampati class 1	2	3	4	Totals	Fitted odds ratio
1	1	5 (6.0)	18 (14.7)	8 (9.7)	1 (1.6)	32	
							1.43
1	2	7 (8.9)	14 (15.1)	11 (7.2)	0 (0.8)	32	
							1.25
2	3	11 (11.0)	16 (14.8)	4 (5.7)	1 (0.5)	32	
							1.09
2	4	13 (11.8)	15 (14.6)	2 (5.1)	2 (0.4)	32	
							0.95
3	5	12 (11.3)	14 (14.7)	5 (5.5)	1 (0.5)	32	
							0.83
3	6	12 (9.5)	12 (15.1)	8 (6.7)	0 (0.7)	32	
							0.73
4	7	5 (6.8)	14 (14.9)	12 (9.0)	1 (1.4)	32	
							0.63
4	8	4 (3.7)	14 (13.2)	11 (12.2)	3 (3.0)	32	
Totals		69	117	61	9	256	

For the observed frequencies for phonation, the odds ratios are different between each adjacent pair of classes: (49/60)/(20/57) = 2.33 between classes 1 and 2, 2.72 between 2 and 3, and 1.35 between 3 and 4. Adding one appropriate term to the minimal equation uses one degree of freedom and "instructs" the model to "choose" (by an iterative process) one, optimum, non-unity odds ratio. In the phonation study this fitted odds ratio is 2.39 (table II) and it leads to fitted frequencies which are a good match to the observed frequencies. This demonstrates that the effect of phonation was largely systematic, as is reflected in the broad similarity of the three observed odds ratios.

Table V shows the likely extreme (lower 95% confidence limit) of the effect of posture on Mallampati class. It was constructed as follows. The 95% confidence limits of the odds ratio were calculated (from the output of GLIM) to be 0.649 and 1.136. The "offset" directive of GLIM was used to combine the more extreme odds ratio (0.649) with the relevant marginal totals of table II, leading to the frequencies given in table V.

The remaining, non-systematic, differences between observed and predicted frequencies in each part of table I may be modelled by adding two more terms to the equation, thereby producing three fitted odds ratios which equal the observed ratios—and a set of predicted frequencies which equal the observed frequencies. If this were applied directly to table II, it would use all the degrees of freedom; but each contingency table in table II is merely a marginal plane of the corresponding three-dimensional contingency table used for the analysis reported in the main text: condition × observer × Mallampati class. To this contingency table it is possible to fit, not only the systematic and non-systematic effects of condition, but also those of the differences between the observers, and those of the interaction between condition and observer, before all the degrees of freedom are used (see below).

The goodness of fit of any model is expressed by its "deviance" in relation to its degrees of freedom. As each explanatory factor is added to the model, the deviance and degrees of freedom are reduced and the significance of the factor may be estimated by referring these changes to a table of χ^2. This process is referred to as an "analysis of deviance" [11] because of the close parallel with analysis of variance.

The analyses of deviance for the models reported in the main text are given in table VI. These confirm the strong systematic effect of phonation: it accounts for 27.5 of the deviance—much more than any other factor. They also confirm, for instance, that the difference between observers was mostly non-systematic in the phonation study (change of 5.4 of a total of 8.9) and entirely so (8.01 of 8.06) in the posture study. Finally, interaction between condition and observer was not significant in either study.

A thorough account of log-linear modelling is given by Agresti [12] and some illuminating examples, using GLIM, are given by Healy [13].

APPENDIX 2

ADDITIONAL FEATURES OF LOG-LINEAR MODELLING USED TO TEST FOR "LEARNING" BY THE OBSERVERS

It was thought that, over such a long series of assessments (128 for each observer in each study), the observers might exhibit a systematic "learning" effect: successive sets of assessments might exhibit a trend towards the better classes (or the worst) as the observers became more practised. Therefore, a serial number (1 to 8) was assigned to each set of 16 assessments made by each observer and this was used as a fourth dimension which was added to the contingency table to be modelled. (The results in the main

TABLE VIII. *Summary analyses of deviance for two studies, using a model which includes serial number of the set of assessments by the observers. Note that the total change of deviance attributed to all the effects of condition and observer and their interaction is identical to that for the simple three-dimensional contingency table used for table VI, although there were some small differences in how the total was shared between the different terms*

	Phonation study					Posture study			
	Residual		Change in				Change in		
Terms in model	df	Deviance	df	Deviance	P	Residual deviance	df	Deviance	P
Minimal model	93	158.65				101.42			
Total of condition, observer and their interaction	84	114.49	9	44.16	< 0.0001	90.34	9	11.08	0.27
"Learning" (common to both observers)									
Linear	83	112.03	1	2.46	0.12	88.75	1	1.59	0.21
Quadratic	82	94.73	1	17.30	< 0.0001	88.39	1	0.36	0.55
Observer–learning interaction (linear + quad.)	80	92.90	2	1.83	0.40	87.42	2	0.96	0.62

Table IX. *Summary analyses of deviance as in table VIII but with ward number fitted before "learning"*

	Phonation study					Posture study			
	Residual		Change in				Change in		
Terms in model	df	Deviance	df	Deviance	*P*	Residual deviance	df	Deviance	*P*
Minimal model	93	158 65				101.42			
Total of condition, observer and their interaction	84	114 49	9	44.16		90.34	9	11.08	0.27
Ward number (all non-systematic)	75	84.27	9	30.22	0 0004	76.00	9	14.34	0.11
"Learning" (common to both observers)									
Linear	74	84.24	1	0 03	0.86	75.74	1	0.26	0.61
Quadratic	73	83.75	1	0 49	0.48	75.08	1	0 66	0.42
Observer–learning interaction (linear + quad.)	71	81 92	2	1 83	0.40	73 50	2	1.58	0 45

text were obtained from a three-dimensional table—condition × observer × Mallampati class.) To allow for the fact that any such learning effect might level out after a time, two systematic terms were assigned to the model: a linear term to fit the overall trend, and a quadratic one (involving the square of the serial number) to allow for the curvature of levelling out.

This produced a marginal contingency table for the phonation study which is shown in table VII. A different odds ratio is fitted to each pair of rows, but the ratios change systematically from row to row: each ratio is 0.874 times the one above. As in tables II and III, each odds ratio applies to all pairs of classes in the relevant pair of rows. This produces a set of "expected" frequencies which make a good match to the observed frequencies. However, rather than showing a trend followed by levelling out, the strong trend towards the better classes in the first half of the study is more than fully reversed in the second half. This seemed an implausible pattern of learning, especially when the corresponding analyses of deviance (table VIII) showed no significant observer–learning interaction (little difference between observers in this "learning" effect) and a negligible learning effect of any kind in the posture study. An alternative explanation is simply that the underlying distribution of Mallampati classes between patients was different in different wards. Inspection of table VII shows that, in classes 1 and 3, where the effect is most noticeable, the frequencies change mainly between wards 1 and 2 and between wards 3 and 4. Therefore, the effect of ward number on Mallampati class was examined by adding it to the model. Then the ward number could fit the difference between wards and the serial number could fit any remaining "learning" effect.

However, there is an important qualitative difference between serial number and ward number: serial number of assessment is an ordinal categorical variable which could plausibly have a systematic effect on the ordinal variable Mallampati class; ward number is not ordinal in this context, only nominal. Wards simply contain different sets of patients who may be expected to exhibit somewhat different distributions of Mallampati classes—an entirely non-systematic effect. Therefore this must be modelled with non-systematic terms, using nine fitted odds ratios (and hence nine degrees of freedom), thereby matching the ratios between the observed frequencies in a ward × Mallampati class plane. The resulting analysis of deviance (table IX) shows that, after incorporating the differences between wards, the systematic "learning" effect disappears: for both the linear and quadratic terms, the reduction in deviance is less than one, and $P > 0.4$.

In contrast, the difference between wards is highly significant ($P = 0.0004$) (table IX). This is surprising. Although the allocation of patients to wards is not a formally randomized process, we were unable to identify any factors which might account for the pattern. We are therefore inclined to regard the low P value simply as signifying a rare chance event.

ACKNOWLEDGEMENT

We are grateful to Dr T. J. Peters, Lecturer in Medical Statistics, University of Wales College of Medicine, for consultations on the interpretation of the results of log–linear modelling.

REFERENCES

1. Solazzi RW, Ward RJ. Analysis of anaesthetic mishaps: the spectrum of medical liability cases. *International Anesthesiology Clinics* 1984; **22** (2): 43–59.
2. Mallampati SR. Clinical signs to predict difficult tracheal intubation (hypothesis). *Canadian Anaesthetists Society Journal* 1983; **30**: 316–317.
3. Mallampati SR, Gatt SP, Gugino LD, Desai SP, Waraksa B, Freiberger D, Liu PL. A clinical sign to predict difficult tracheal intubation: a prospective study. *Canadian Anaesthetists Society Journal* 1985; **32**: 429–434.
4. Samsoon GLT, Young JRB. Difficult tracheal intubation: a retrospective study. *Anaesthesia* 1987; **42**: 487–490.
5. Armitage P, Berry G. *Statistical Methods in Medical Research*, 2nd Edn. Oxford: Blackwell Scientific, 1987; 239–245.
6. Nelder JA. Log linear models for contingency tables: a generalization of classical least squares. *Applied Statistics* 1974; **23**: 323–329.
7. Wilson ME, John R. Problems with the Mallampati sign. *Anaesthesia* 1990; **45**: 486–487.
8. Oates JDL, Oates PD, Pearsall EJ, McLeod AD, Howie JC. Phonation affects Mallampati class. *Anaesthesia* 1990; **45**: 984.
9. Oates JDL, Macleod AD, Oates PD, Pearsall FJ, Howie JC, Murray GD. Comparison of two methods for predicting difficult intubation. *British Journal of Anaesthesia* 1991; **66**: 305–309.
10. Turnbull A, Tindall VR, Beard RW, Robson G, Dawson IMP, Cloake EP, Ashley JSA, Botting B. *Report on Confidential Enquiries into Maternal Deaths in England and Wales 1982–1984*. London: HMSO, 1989; 99–101.
11. Payne CD. *The GLIM System Release 3.77: Manual*. Oxford: Numerical Algorithms Group, 1987; 107.
12. Agresti A. *Analysis of Ordinal Categorical Data*. New York: Wiley, 1984; 14.
13. Healy MJR. *GLIM: an Introduction*. Oxford: Clarendon Press, 1988; 94–99.