On the relationship between the efficiency and the quality of the consultation. A validity study

Jo Goedhuys and Jan-Joost Rethansa


Objectives. The aim of this study was to determine how the medical performance of physicians during consultations is related to doctor–patient communication and satisfaction of patients, taking into account the actual length of the consultations. In addition, we studied the validity of the ‘efficiency-per-time score’ as a measure of competence.

Methods. General practice trainees participated in a test situation in which they were confronted with six consultations with standardized (simulated) patients (SPs). All consultations were videotaped and evaluated by multiple observers, using national guidelines on medical content and on communication. The SPs scored satisfaction with the consultation using a satisfaction checklist. Forty GP-trainees were invited, of whom 34 participated. The main outcome measures were the number of obligatory actions undertaken by the GP-trainees, total number of actions undertaken, consultation time, efficiency-per-time score, patient satisfaction and quality of communication score, and the Pearson correlations between these measures.

Results. There was a negative correlation between the ‘efficiency-per-time score’ of the GP-trainees and the satisfaction of the SPs in five of the six consultations [Pearson r from –0.29 (P < 0.05) to –0.58 (P < 0.001)] and between the ‘efficiency-per-time score’ and the quality of the communication in three of the six consultations [Pearson r from –0.34 (P < 0.05) to –0.51 (P < 0.001)].

Conclusions. Short consultations with high technical medical efficiency seem to be related to bad communication and dissatisfied patients, thus questioning the validity of the ‘efficiency-per-time score’ as a measure of competence.

Keywords. Consultation, competence, doctor–patient communication, patient satisfaction, performance

Introduction

‘Competence’ has been defined as “what a doctor is capable of doing” and ‘performance’ as “what the doctor actually does in daily practice”. Most licensing bodies and medical schools base their decision about which candidates will or will not qualify as practising doctors by tests to be taken by candidates under real examination conditions, thus assessing the competence of candidates. Thereby it is assumed that competence is a predictor of performance. Recent studies, however, have shown that the relationship between competence and performance is not as straightforward as might be expected.

In studies where GPs were confronted with the same (standardized) patients (SPs) both within actual practice (incognito patients) and within a test situation, the correlation between performance and competence was nearly zero for all variables used, except for the length of the consultation and for the ‘efficiency,’ which showed Pearson correlations up to 0.49. Efficiency was defined as ‘the number of obligatory actions divided by the total number of actions undertaken’.

The use of the consultation time in the relationship between competence and performance originates from literature on medical problem solving, suggesting that the amount of time a doctor needs to solve a medical problem is a reflection of his or her expertise. Because expert doctors show better and more adequate processing of relevant patient information than less experienced colleagues within a certain time span, it was concluded in previous research that time combined with efficiency is an important variable to estimate expertise level. Indeed, in the comparison between competence and performance by means of SPs, this combination of length of
consultation with efficiency level produced a Pearson correlation of 0.45 (P < 0.01): the ‘efficiency-per-time score’.5 This result suggested that it is possible to identify a ‘quality factor’ that remains stable across different situations, making it possible to classify physicians on the basis of a single, although comprehensive, index. Recent research, however, has shown that even between repeated measurements in performance settings, there is no such stability.10

Other researchers came to different conclusions about the role of time in the estimation of expertise and the quality of care delivered. Hofmans-Okkes,11 after an extensive literature review, concluded that the role of time remained unclear, whereas Howie et al.,12 found a positive relationship between patient satisfaction and consultation length.

To test the validity of the ‘efficiency-per-time score’ further, it was decided to explore to what extent this score relates to other indices of quality of care, such as the quality of the communication and the patient satisfaction. The latter variables (quality of communication and patient satisfaction) have been studied extensively13,14 and are known to be related to the quality of care delivered.

Subjects and methods

We invited 40 general practice trainees to participate in a test situation in which they were each confronted with six consultations with SPs. All consultations were videotaped and evaluated by multiple observers, using national guidelines on medical content and on communication. The SPs scored satisfaction with the consultation using a satisfaction checklist. Further details of this procedure are set out in the following paragraphs.

Case selection and development of instruments

Cases were selected by the following criteria: they should have a high prevalence in general practice, have varying degrees of features of both medical-technical and psychosocial content, and have the possibility to be simulated by SPs. For reasons of feasibility, six cases were selected: shoulder pain, sore throat (both male roles), urinary tract infection, contraception, hyperventilation and fatigue (female roles). The latter three cases contained more psychosocial aspects than the first three cases. The SPs (four per case) were selected and trained according to standard procedures.15–17 After 8 hours of training, they were able to present their knowledge in a standardized way and to use the EVA.PAT satisfaction checklist (see below).

Next, checklists were developed to evaluate the quality of the medical care provided in actual consultations with these six cases. Four of these checklists were based on the National Guideline Programme of the Dutch College of General Practitioners.18 The other two (hyperventilation and fatigue) were designed by a team of Flemish experts in general practice. All actions mentioned in these guidelines were given one of three qualifications: ‘obligatory’, ‘intermediate’ or ‘superfluous’. To validate these checklists, the opinion of 20 GPs—teaching at one of the four Dutch Universities in Belgium—was sought. According to their advice, some changes were introduced in the checklists.

To evaluate the quality of the communication, the MAAS-Global (Maastricht Anamnesis and Advice Scorelist) was chosen.19 This instrument for assessing the quality of doctor–patient communication consists of 11 global items to be scored from 1 to 7, yielding a total score of between 11 and 77. It has proven to be reliable (Cronbach’s alpha >0.80).

To measure the satisfaction of the SPs after each consultation, a questionnaire with seven items (EVA.PAT) was developed. It was based on questionnaires used and validated in similar research.20,21

Competence assessment

For reasons of feasibility, second year trainees in general practice were chosen as subjects for this study. (In Belgium, vocational training for GPs takes 3 years. The second and third year consist mainly of working as a GP in a GP trainers’ practice, under his supervision. Trainees know they can be asked to participate in research as a part of their training.) To guarantee that the sample would be representative for other second year GP-trainees (n = 184), trainees were stratified according to gender, age and university of graduation. The resulting sample of 40 trainees was invited to come to a test situation at our medical school, where the trainees conducted consultations with six SPs. For this purpose, each trainee had his own consultation room, a waiting room with the six SPs and 2 hours time, so there was virtually no time constraint. All consultations were videotaped to allow a detailed observation afterwards. After each consultation, the SPs filled in the satisfaction questionnaire.

Scores calculated

To evaluate the videotaped consultations, seven GPs were trained (15 hours) to use the medical checklists and the MAAS-Global. To avoid bias of observer’s evaluation of one consultation of a trainee on other consultations of the same trainee (the ‘halo’ effect), the observers received a videotape with 40 consultations in a fixed order, making sure they observed no more than two consultations of the same trainee. To calculate inter-observer consistency, 20 consultations per case were evaluated by two or more observers.

For each consultation, the following scores were calculated from the observation of the videotapes: the number of obligatory actions performed, the total number of actions performed in each consultation, the...
efficiency score (the ratio of the number of obligatory actions performed divided by the total number of actions performed), the consultation time in seconds, the efficiency-per-time score (the efficiency score divided by the length of the consultation) and the MAAS score. The EVA.PAT scores were calculated from the satisfaction questionnaires.

Reliability of the instruments
To compute the scores for the number of obligatory, intermediate and superfluous actions undertaken, only reliable items (actions) were taken into account. Reliability on the item level was defined as an inter-observer agreement of $\kappa > 0.60$, but others suffered from the ‘base rate problem’, i.e. very skew marginal distributions, indicating a very high or low prevalence of the action under study. In these cases, kappa values cannot be interpreted properly.23

After this selection of items, reliability of the number of obligatory, intermediate and superfluous actions undertaken was calculated as the Pearson correlation between the numbers obtained by two observers.

For the MAAS, the reliability was established again using the Cronbach’s $\alpha$-coefficient. For the EVA.PAT questionnaire, reliability was also analysed using the $\alpha$-coefficient.

Results
Subjects
Of the 40 invited GP-trainees, 34 (85%) participated. These 34 trainees proved to be representative for the population of trainees in general practice in Belgium in terms of (i) membership of the four universities, (ii) gender, (iii) age and (iv) academic results. The results concerning the first two variables are shown in Table 1. The trainees who were absent, were so for practical reasons, and were not different from the participating trainees on these four variables.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>University and gender of the participating trainees</th>
</tr>
</thead>
<tbody>
<tr>
<td>KUL</td>
<td>RUG</td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
</tr>
<tr>
<td>Female</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
</tr>
</tbody>
</table>

KUL, Katolieke Universiteit Leuven; RUG, Universiteit Ghent; UIA, Universiteit Antwerpen; VUB, Vrije Universiteit Brussel.

Reliability
The selection of items of the medical checklists resulted in a loss of 62 of the 316 items, i.e. 19% had an inter-observer agreement of $\kappa > 0.80$. Other scores reported in this article were computed using the remaining reliable items only.

Inter-observer reliability ranged from $r = 0.73$ to $r = 0.87$ for the number of obligatory actions observed, and from $r = 0.71$ to $r = 0.90$ for the total number of actions observed (all Pearson correlations). Thereby the inter-observer reliability of the efficiency-per-time score ranges from $r = 0.79$ to $r = 0.97$.

Reliability analysis of the EVA.PAT questionnaire showed $\alpha$-coefficients $>0.80$ for the six cases (0.84–0.95).

The generalizability coefficient for the MAAS instrument was $\alpha = 0.80$.

Relationships between the variables
Table 2 shows the observed Pearson correlations between the variables.

For four cases (urinary tract infection, contraception, hyperventilation and fatigue; those with more psychosocial aspects) we found significant correlations between all variables, except one. Longer consultations tended to be evaluated better by the SPs in terms of satisfaction (Pearson correlations from 0.29 to 0.63). Longer consultations were also rated higher for the quality of the communication by the GP-observers (Pearson correlations from 0.29 to 0.61). Thereby the quality of

Table 2 | Intercorrelations (Pearson $r$, observed correlations) between consultation length (time), satisfaction of the standardized patient (EVA.PAT), quality of the communication (MAAS) and the efficiency-per-time score (Effic/time) for six cases (34 GPs)

<table>
<thead>
<tr>
<th>Case</th>
<th>Time EVA.PAT</th>
<th>Time MAAS</th>
<th>EVA.PAT MAAS</th>
<th>Effic/time MAAS</th>
<th>Effic/time EVA.PAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder</td>
<td>0.04</td>
<td>-0.15</td>
<td>0.14</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>Throat</td>
<td>0.24</td>
<td>0.07</td>
<td>0.36*</td>
<td>-0.08</td>
<td>-0.29*</td>
</tr>
<tr>
<td>Urinary</td>
<td>0.50**</td>
<td>0.61***</td>
<td>0.62***</td>
<td>-0.51***</td>
<td>-0.47**</td>
</tr>
<tr>
<td>Contraception</td>
<td>0.34*</td>
<td>0.29*</td>
<td>0.53***</td>
<td>-0.34*</td>
<td>-0.49**</td>
</tr>
<tr>
<td>Hyperventilation</td>
<td>0.29*</td>
<td>0.29*</td>
<td>0.64***</td>
<td>-0.23</td>
<td>-0.43**</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.63***</td>
<td>0.47**</td>
<td>0.55***</td>
<td>-0.41**</td>
<td>-0.58***</td>
</tr>
</tbody>
</table>

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. 

---

594 Family Practice—an international journal
the communication and the satisfaction of the patient are also positively correlated (Pearson correlations from 0.53 to 0.64). Since the consultation length is in the denominator of the efficiency-per-time score, this efficiency-per-time score is negatively correlated with satisfaction (Pearson correlations from –0.43 to –0.58) and quality of communication (Pearson correlations from –0.23 to –0.51).

For the two more medical–technical-oriented consultations (shoulder and sore throat), only two out of 10 correlations were significant: between patient satisfaction and quality of communication ($r = 0.36$) and between patient satisfaction and the efficiency-per-time score ($r = –0.29$), both for the sore throat case.

Discussion

Since the efficiency-per-time score consists of the variables ‘time’ and ‘actions undertaken’, we discuss them separately.

Concerning the variable time, we observed important correlations between time, satisfaction and communication. It is worthwhile remembering that these three variables were all measured in a statistically independent way, i.e. by different observers (stopwatch, standardized patient and trained observer).

The correlation between the satisfaction of the SP and the quality of the communication as observed by trained GPs suggests that our SPs included an appreciation of the communication in their total impression of satisfaction. Furthermore, it seems to indicate that our SPs used similar criteria to evaluate consultations as the criteria underlying the MAAS instrument. These criteria are considered of crucial importance for good medical care since the last decades: taking into account the ideas, concerns and expectations of the patient, explaining medical findings in an understandable way, negotiating aspects of the management plan with the patient, making the decision process transparent, etc. It is possible that the correlations emerged in our research because most of our SPs were university students: perhaps they are very susceptible to ‘measure’ these values. Also, it is possible that our results are typical for GP-trainees and would not be replicated with experienced GPs.

With these data, it is not possible to make inferences about the causal relationship between the three variables. We hypothesize however, that the ‘pure’ length of the consultation is not the determinant of the quality of the consultation or of the satisfaction of the SP. Rather, we feel that following the principles of good communication, thereby realizing the values mentioned above, inevitably takes more time and brings more satisfaction. Careful listening, explaining and negotiating all take time that cannot be compressed unless at the cost of the quality of the job. Thus, using little time may be an indication of expertise, but is seemingly not compatible with good communication or with satisfied patients.

The second variable that constitutes the efficiency-per-time score concerns the number of ‘obligatory’ actions undertaken. Of course, caution is required when using the word ‘obligatory’. Guidelines do change and may be different in different countries. Therefore, the validity of the efficiency score as an operationalization of expertise may be restricted in time and space.

Our conclusion is that the efficiency-per-time score may have been useful in exploring the relationship between competence and performance, but is probably not a good indicator of the quality of the consultation. Performing a lot of correct medical actions within a short period of time may well be an indication of the technical expertise of the GP, but excludes another essential aspect of good medical care: taking time to take the patient into account as a person.

Acknowledgements

JG is grateful to Kate Billingham for her inspiring comments.

References


