

Experience of using rating scales for the assessment of vocational trainees in general practice

F DIFFORD

R C W HUGHES

SUMMARY. For two years trainers in the northern half of the south western region have been asked to assess their trainees with a condensed version of the new Manchester rating scales for vocational training in general practice and to send copies to the regional adviser. The condensed version retains the 23 scales but does not include the subscales. Trainers' workshops and comments from individuals suggested that no radical improvements to the scales were required and that they were an adequate statement of what was required from trainees. The response rate of 89% for returning at least one assessment and 38% for returning all three suggested that use of this condensed version would be more feasible than using the full version of the scales. Analysis of 86 first assessments and 48 sets of three assessments showed a range of variation both in trainee ratings and use of individual scales that was consistent with expectations. A lack of significant rank correlations between the scales indicated that each of the scales measured a different characteristic of behaviour.

We suggest that the most useful way to achieve systematic assessment of vocational trainees is by the use of the 23 main rating scales, with regional variations where appropriate.

Introduction

IN 1987 the Joint Committee on Postgraduate Training in General Practice and its parent bodies accepted the report of a joint working party on general practitioner trainee assessment.¹ This included a recommendation for the use of rating scales and suggested that regional advisers establish a system to manage certification based on the new extended Manchester rating scales² in order to achieve the greatest possible consistency among the raters.

We felt that the use of these scales took a disproportionate amount of training time and therefore devised a condensed version (Appendix 1), retaining all 23 scales but removing the subscales. The trainer is asked to score the trainee on a scale from 10 (outstanding), through five (average) to one (very poor) for each of the 23 criteria, having considered how he or she performs on a number of substatements without actually rating them. The retention of all 23 scales made it possible to continue to use the complete scales in cases of doubt but it should be noted that no advice has been provided on converting the frequency rating on the subscales into a number on the main scale.

These condensed scales were introduced to Gloucestershire,

Avon and Somerset trainers in 1987 and from August 1988 trainers were asked to send copies of completed scales to the regional adviser in the fourth, eighth and 12 months of training. The scales were not recommended for initial assessment because the trainer would not have had the opportunity to observe the trainee and the trainee would not have had time to adjust to general practice. It is emphasized that the scales are not a method of assessment in themselves, but a way of bringing together the results of other assessments, for example analysis of consultation videotapes, sitting-in on consultations and random case analysis.

The aim was to encourage the use of the scales in both formative and summative assessment and to provide feedback to promote consistency, particularly among trainees scoring at the lower end of the scales where competency may be in question. Formative assessment must be shared with the trainee, and his or her relative scoring in the different scales can show strengths and weaknesses to be worked upon. Trainers were not encouraged to total the 23 scales since different weightings would attach to each. Summative assessment can be used not only to see if the trainer and trainee achieve their aims but also to support certification of competence to practise by showing that acceptable scores have been achieved. One rating of two or several scores of three were suggested as indications for not certificating the trainee as competent.

This paper examines the results of using the scales for two years in order to determine their feasibility, reliability and validity.

Method

An assessment pack containing a variety of materials was sent to each trainer in the northern half of the south western region when his or her trainee started with the practice. In addition, a reminder was sent at the appropriate assessment time and if necessary after the trainees had completed their trainee year. The system was one of exhortation only. A database of trainees and trainers held in the regional adviser's office was used to record training details, including the fact that assessments had been returned, but not the scores, which were kept on a separate database not accessed for routine enquiries.

Trainers' comments on the scales were invited at nine workshops and individually. They were also invited to append two additional scales to the 23 on each assessment form.

The returned scales were analysed using means and spread of results looking for variation in the performance of trainees over the three assessments and between individual trainees on each assessment. Variation in the performance of trainees as a group on individual scales over the three assessments and between scales was also analysed using means and spread. The correlation of ratings of trainees on different scales was examined using Spearman's rank correlation coefficient.

Results

The trainers' workshops, convened to discuss the scales, always agreed that they were a good statement of what was expected from trainees and did not call for radical changes or additions. Preventive medicine, practice management and ethics were

F Difford, FRCGP, associate advisor in general practice and R C W Hughes, FRCGP, regional advisor in general practice, Medical Postgraduate Department, University of Bristol.
Submitted: 17 December 1990; accepted: 5 April 1991.

© *British Journal of General Practice*, 1991, 41, 360-364.

suggested as additions by some but the behaviour required in these areas was covered by other scales and to include them could lead to many scales where the main content difference would be knowledge based, and best measured in other ways. Fourteen trainers used the extra scales on the assessment forms, but these scales were either qualities of individual trainees such as enthusiasm, personal organization, personability, flexibility and self confidence, which were less subject to objective assessment or were covered by other scales. One trainer felt that the scales were too much concerned with what the trainee did rather than felt.

The response of trainers to the assessment plan is shown in Table 1. After two years complete sets of three rating scales were returned for 48 trainees and a further 86 first assessments were available for analysis, eight having been excluded because the trainers stated that they had changed their rating method or because there were mistakes on the form.

Table 1. Use of rating scales by the trainers over the two study years.

	No. (%) of trainers	
	Aug 88 -Jul 89 (n = 80)	Aug 89 -Jul 90 (n = 92)
Returning a single assessment	12 (15)	23 (25)
Returning two assessments	13 (16)	24 (26)
Returning three assessments	15 (19)	35 (38)
Total returning at least one assessment	40 (50)	82 (89)

n = total number of trainers with trainees who completed training in this year.

The 96 trainers had been trainers for a mean of 6.1 years of which a mean of 3.9 years had been spent with trainees.

A dot plot of the mean rating scale score for each of the 134 trainees at the first assessment is shown in Figure 1. The overall mean score for the 134 trainees at the first assessment was 5.37. Figure 2 shows the number of scale points used for each trainee at the first assessment on all 23 scales. Most trainers used only three or four points overall showing fairly uniform ability across the scales for these trainees. A dot plot of the change in mean rating scale score from the first to the third assessment for the 48 trainees where three complete assessments were obtained is shown in Figure 3. The mean change during the year was +1.24 (from 5.39 to 6.63) — 0.74 between the first and second assessments and 0.50 between the second and third. Sixteen trainees showed nearly all the gain in the first eight months and six showed the majority in the last four months.

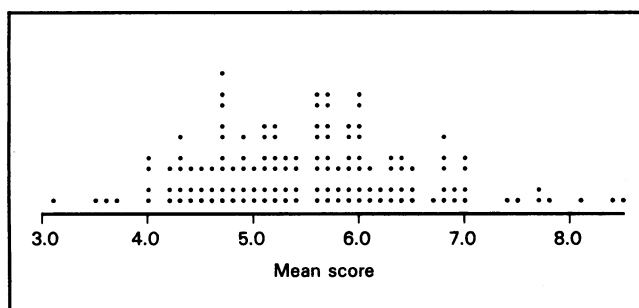


Figure 1. Mean rating score on the 23 scales for 134 trainees at the first assessment.

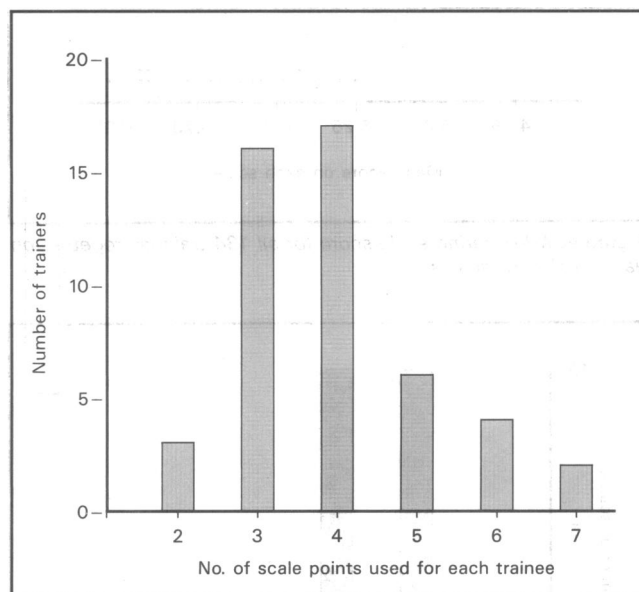


Figure 2. Number of scale points used by 48 trainers for each trainee at the first assessment.

Figure 4 shows the mean rating scale score of the group of 134 trainees as a whole on each of the 23 scales at the first assessment. The highest scores were obtained for consideration and specific skills in physical examination (items 4 and 6, Appendix 1), availability (item 19), involvement (20), working with colleagues (22) and all aspects of emergency care (16–18). The lowest score was obtained for using community resources (item 12). Figure 5 shows the number of consecutive points used by the trainers when grading the 134 trainees on each of the 23 scales; this varied between five and eight points. Technical aspects of prescribing (item 13, Appendix 1) showed the least difference with scale points ranging from four to eight while availability and range of emergency situations (items 19 and 18) showed the greatest difference at three to 10 scale points. Figure 6 shows the change in mean rating scale scores from the first to the third assessments for the 48 trainees together for each of the 23 scales. The biggest change was found for using community resources (item 12, Appendix 1) followed by general and special skills in history taking (1 and 2). The smallest improvements were seen in general observation, general approach and specific skills in physical examination (items 3, 5 and 6), communication (21) and coping with uncertainty (11).

There was no significant rank correlation between the 23 scales.

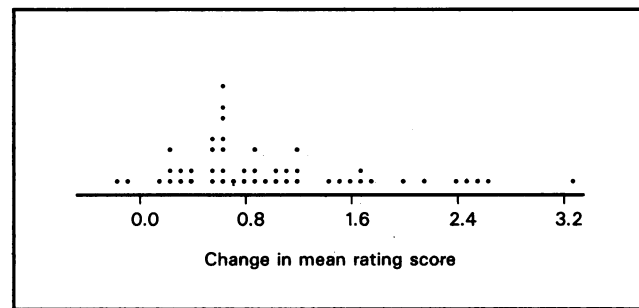


Figure 3. Change in mean rating score between the first and third assessments for 48 trainees.

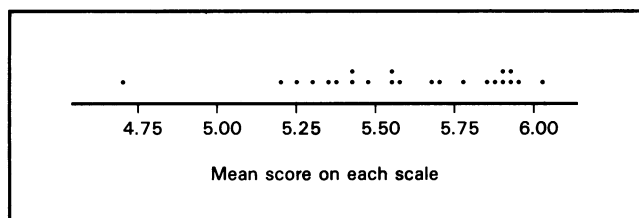


Figure 4. Mean rating scale score for all 134 trainees together on each of the 23 scales.

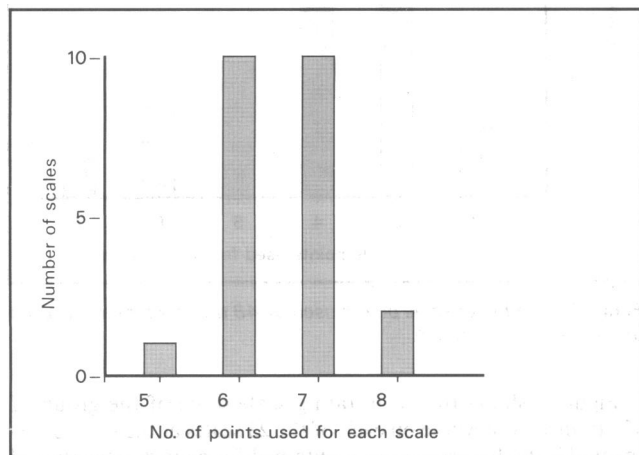


Figure 5. Number of consecutive points used by the trainers when grading the 134 trainees on each of the 23 scales.

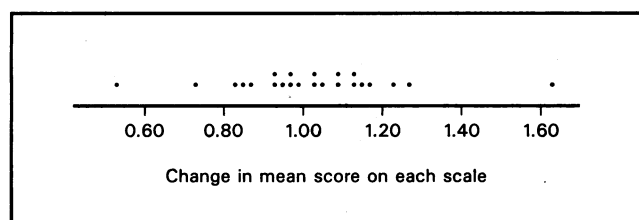


Figure 6. Change in mean rating score between first and third assessments for all 48 trainees together on each of the 23 scales.

Discussion

Validity

There is a wide variety in the content and emphasis of the work of general practitioners. This variety and the constant evolution required by society and medical advances make it difficult to define precisely how we should measure competence. The general practitioner trainee rating scales would seem to offer sufficient detail over a broad area but the only basis for their validity is that they were the result of a considerable amount of work done by an experienced and geographically spread group of general practitioner educationalists.

Our findings can only offer further limited support for the validity of these scales, and it must be remembered that they are not used to show degrees of excellence but only to clarify the possible reasons for lack of competence.

The lack of significant rank correlation between the scales suggests that each measures a different attribute and that none should be omitted.

The next step in testing validity would be to send a questionnaire to principals and those in other careers who had taken part in assessment using the scales during their traineeship three years

previously. Those in practice could be asked to consider the relevance of the scales to their work while all could comment on the meaning and influence of their scores on their subsequent professional development. Prior consent would be required to keep the scales and to contact the ex-trainee.

Reliability

As for the validity, the reliability of these scales is only important in the context used, especially that of assessing minimum standards of competence. We can only show evidence for the reliability of the scales by demonstrating that the results are consistent with expectations.

Measuring trainee performance. The overall mean score of 5.37 at four months was a little high when the average for a general practitioner was regarded as five but well below the ubiquitous seven out of 10. On some scales the trainee might be expected to do better than an average general practitioner because of the additional time available in consultations. The variation in the mean scores of individual trainees (Figure 1) shows several trainees with very high scores. This might be the result of the trainer rating incorrectly rather than of trainee ability, although some trainers commented on their exceptional trainee and some trainers giving high scores rated other trainees in the more usual range. The variation in the number of points on the scales used by trainers at the first assessment (Figure 2) was compatible with some trainees showing fairly uniform ability across the scales and others showing a wide range of strengths and weaknesses.

The improvement in mean score between the first and third assessments of 1.24 was credible in doctors going through at least their fourth year of postgraduate training. The greater improvement shown overall from four to eight months compared with eight to 12 months agrees with the experience of trainers. So too does the variation in timing of improvement in individual trainees — some showing most improvement between first and second assessments and others showing most in the last four months. A few trainees showed no improvement in rating and others a remarkable gain (Figure 3) but all appeared to be within the bounds of possibility.

Group performance on scales. Looking at the way in which the trainees performed as a group on the different scales at the first assessment (Figure 4) there was an expected variation, the extremes of which could be explained by the skills brought to general practice training from previous experience. Thus the high scores for specific skills and considerateness in physical examination could be ascribed to the focus on physical examination from medical student days, those for availability, involvement and working with colleagues to the trainees' part in a hospital team, and those for all aspects of emergency care to the considerable part of their career spent on call. Similarly the low score for using community resources reflects the lack of opportunity for such usage in hospital medicine.

The number of points used for each of the rating scales for the group of trainees varied from five to eight points (Figure 5), reflecting either a different range of ability of trainees in each scale or the degree of confidence in scoring the scale owing to variations in opportunity to observe or faults in the scales themselves. Thus technical aspects of prescribing showed least difference ranging from four to eight, perhaps reflecting less variation in ability while availability and range of emergency situations showed the greatest difference, three to 10, these being areas producing obvious effects in the practice, which the trainer can score.

The improvement achieved by the trainees as a group varied between scales (Figure 6). The big increase in scoring for using

community resources is accounted for by low initial scores while the small increase for specific skills in physical examination can be related to the high initial scores. History taking, both general and special skills, improved considerably; this is a skill at the very heart of general practice work. However, communication and coping with uncertainty showed less improvement; these are areas in which personality may play a greater part. General observation and general approach in physical examination also showed less improvement, perhaps because trainers find these aspects difficult to measure.

While none of these findings can have statistical significance they are consistent with the majority of trainers using the scales carefully.

Feasibility

The high response rate achieved in the second year for a voluntary system suggests that trainers could cope well with these assessments but that further cooperation is unlikely without more direction. A sizeable minority of trainers, including some of those who returned scales, expressed doubts about assessment using this amount of detail or with this level of quantitative assessment. Since the full version of the scales contains an additional 166 subscales it seems unlikely that an acceptable take-up rate of the full scales could be achieved in this region.

Unlike teaching and the supervision of training which can be shared between trainers and partners in the practice the assessment of satisfactory completion of training is the exclusive responsibility of the named trainer. Currently, the average trainer in our half region has had trainees for approximately four years. This means he or she has only a limited opportunity to become familiar with the use of rating scales and even less to develop his or her own comprehensive system of assessment. This makes it important that trainer workshops should regularly discuss assessment.

There is an increasing call for the vocational training regulations form one (VTR1) to certificate competence to practise rather than just 'satisfactory completion' of a period of training.³ An equally systematic but voluntary system of assessment by returning modified essay question and multiple choice question papers to the regional adviser's office for marking has been in use in this half region for the last three years but has an uptake rate of only 54% (the degree of supervision, irregular timing of participation and low uptake of this scheme have not provided reliable data for correlation with rating scales).

We therefore suggest that the most useful way to achieve systematic assessment is by the use of all 23 main rating scales with regional variation depending on local circumstances.

Appendix 1. A condensed version of the general practitioner trainee rating scales. These scales are Crown copyright and are reproduced with the permission of the controller of HMSO.

1. History taking: general

He or she is skilful in acquiring information about the patient. He is attentive and willing to take time to listen without interruption. He allows the patient to express his or her own ideas and concerns. He clarifies the reason for the patient's visit and the patient's expectations.

2. History taking: special skills

He is skilful in gathering information about psychiatric symptoms. He follows up psychological and social factors where appropriate. He demonstrates skill in discussing sensitive and personal matters.

3. Physical examination. 1: general observation

He shows that he has taken note of the patient's appearance, behaviour and physical activity. Where appropriate, he increases information from a visit to the home environment.

He shows that he can use these sorts of information in understanding the patient.

4. Physical examination. 2: considerateness

He is sensitive to the patient's needs and feelings, especially privacy. He helps where necessary and continues to listen while examining. He explains to the patient what he is doing and why.

5. Physical examination. 3: general approach

He orders the examination selectively and thoroughly, reflecting the history established. He recognizes when to carry out an extended examination. He rechecks doubtful findings.

6. Physical examination. 4: specific skills

He examines each system and each organ properly and detects physical signs adequately. He can demonstrate the correct technique of using auroscope, ophthalmoscope, sphygmomanometer, stethoscope, patella hammer, tuning fork, vaginal speculum, proctoscope and peak flow meter.

7. Problem definition. 1: hypothesis formation

His initial ideas about a problem include all the common and important causes. His hypotheses incorporate unexplained findings or apparent inconsistencies.

8. Problem definition. 2: hypothesis testing

He tests his hypothesis on a sound estimate of frequency and probability. He uses sound clinical judgement to assess the degree of seriousness. He moves logically from rejecting one hypothesis to considering the next. He ends the consultation only when the problem has been adequately defined.

9. Problem definition. 3: coping with complexity

He incorporates into the analysis of the problem what he knows of the patient's life and background, so that he defines it in physical, psychological and social terms. He includes causal factors, for example smoking; contributory factors, for example obesity; and associated factors, for example job loss. He takes into account other diseases and problems the patient is known to have or to be at risk of.

10. Problem definition. 4: practicability

He defines the problem in terms of facts he feels sure of. He makes it clear whether further information is required. He formulates a safe, effective and acceptable plan of management.

11. Management. 1: coping with uncertainty

Having excluded immediate risks, he is prepared to watch and wait to allow things to become clear, while gaining the patient's trust. He uses diagnostic facilities appropriately and economically and with regard to the patient's feelings and convenience. Where necessary, he gets consultant help in the most effective way.

12. Management. 2: using community resources

He knows the full range of services available in the community. He refers appropriately after obtaining the patient's agreement. He encourages patients to join self-help and support groups.

13. Management. 3: prescribing — technical

He prescribes the most appropriate drug in suitable quantity only after getting adequate information. He prescribes only after consideration of all forms of management. He takes due regard of hypersensitivities, drug interaction and side effects.

14. Management. 4: prescribing — interpersonal

He involves the patient in defining the aims of treatment. He advises the patient on dosage, duration of treatment, side effects and specific precautions. He arranges for feedback for effectiveness and side effects. He considers the cost of prescriptions. He makes sound decisions about starting repeat prescribing. He uses proper methods of monitoring repeat prescriptions.

15. Medical records

He reads the records before the consultation and uses them in clinical decision making.

He writes records promptly without hindering the interaction.

His records are complete, concise, legible and always available to the practice.

He uses other components of the medical record system.

16. Emergency care. 1: initial assessment

In an emergency, he puts the patient's needs above his own convenience. He uses time with the messenger to obtain relevant information and give interim advice and comfort.

He decides correctly on the degree of urgency.

He takes the right equipment.

17. Emergency care. 2: management

He takes appropriate responsibility at the scene of an emergency.

He makes an accurate assessment using information from the patient and witnesses.

He uses drugs, equipment and emergency resources properly.

He handles his own emotions well enough to remain controlled and comfort and reassure the patient and helpers.

18. Emergency care. 3: range of situations

He can cope to the extent of preserving life and executing an initial plan for the common emergencies seen in general practice: haemorrhage, coma, trauma, abdominal pain, respiratory distress, fits, chest pain and psychiatric problems.

19. Professionalism. 1: availability

He is punctual and uses consulting time flexibly.

He accepts and adjusts to variations in workload.

He is openly available and patients consult him readily.

20. Professionalism. 2: involvement

He presents himself appropriately for work.

He shows respect for the patient's customs, values, ideas and attitudes.

He demonstrates to patients an interest in and concern for their family and work situations.

He makes plans which take a patient's personal situation into account.

He shows tenacity in helping patients in difficult or frustrating situations.

He encourages them to be more self-aware, questioning and self-reliant.

He gives health education when the opportunity or need arises.

21. Professionalism. 3: communication

He communicates effectively with a wide range of patients, using language with which they are comfortable and is thus able to comfort and support them.

He communicates well with small children, adolescent males, adolescent females, the mentally handicapped, the old, the socially disadvantaged and those in ethnic minorities; and on occasion with patients who are aggressive, angry, deceptive, distressed, disbelieving or flirtatious; and with more than one patient at a time.

22. Professionalism. 4: working with colleagues

He shows by his behaviour towards other members of the primary care team and practice that he understands their role and responsibilities.

He values their help and respects their professional autonomy.

He balances the patient's right to confidentiality with a colleague's need for information.

He respects a colleague's right to confidentiality.

He works well with principals in practice, practice employed nurses, district nurses, associated social workers, practice administrative staff, health visitors, midwives and other health care professionals.

23. Personal development

He continues to learn by reading books and journals and making good use of postgraduate meetings.

He can by self-evaluation or peer review identify gaps in his competence.

He undertakes audit into such areas as prescribing, referral follow-up, consultation skills and effect of disposition on work.

References

1. Joint Committee on Postgraduate Training for General Practice. *Assessment and vocational training for general practice. Final report of a JCPTGP working party.* London: JCPTGP, 1987.
2. Centre for Primary Care Research, Department of General Practice, University of Manchester. *Rating scales for vocational training in general practice. Occasional paper 40.* London: Royal College of General Practitioners, 1988.
3. Joint Committee on Postgraduate Training for General Practice, the Royal College of General Practitioners and the General Medical Services Committee. *Vocational training regulations: the meaning of 'satisfactory completion'.* London: JCPTGP, 1990.

Acknowledgements

We are grateful for the statistical advice of Mr A Hughes and for the cooperation and helpful suggestions of trainers in Gloucestershire, Avon and Somerset.

Address for correspondence

Dr F Difford, Medical Postgraduate Department, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR.

RCGP

Scientific
Foundation
Board

**RESEARCH
FUNDING**

Applications are now being received for grants for research in or relating to general medical practice, for consideration at the November 1991 meeting of the Scientific Foundation Board. In addition to its general fund the Board also administers specific

funds including the Windebank Fund for research into diabetes.

The Scientific Foundation Board's definition of research is catholic and includes educational research, observational as well as experimental studies, and accepts the methodologies of social science as valid. It is not in a position to fund educational activities.

If the study involves any intervention or raises issues of confidentiality it is wise to obtain advance approval from an appropriate research ethics committee otherwise a decision to award a grant may be conditional upon such approval.

Studies which do not, in the opinion of the Board, offer a reasonable chance of answering the question posed will be rejected. It may sometimes be useful to seek expert advice on protocol design before submitting an application.

Care should be taken to ensure that costs are accurately forecast and that matters such as inflation and salary increases are included.

The annual sum of money available is not large by absolute standards and grant applications for sums in excess of £15 000 are unlikely to be considered.

Application forms are obtainable from the Clerk of the Board at: The Scientific Foundation Board, 14 Princes Gate, London SW7 1PU. *The closing date for receipt of completed applications is 27 September 1991; any forms received after that date will, unfortunately, be ineligible for consideration.*