Factors influencing diagnostic decision-making

Kathleen SN Callaghan PhD

Human Factors Group, Department of Surgery, Faculty of Medical and Health Sciences, The University of Auckland, New Zealand

ABSTRACT

INTRODUCTION: Identifying influences on diagnostic decisions is important because diagnostic errors often have far-reaching consequences for an individual's future within the workforce and their eligibility for Accident Compensation Corporation–funded treatment. Most investigations of factors biasing decision making have used quantitative techniques rather than qualitative methods.

AIM: To identify factors influencing GPs' diagnostic decision-making and to develop a valid questionnaire to determine the desirability and importance of each factor's influence.

METHODS: Focus groups and the Delphi method were combined with Rasch analysis to identify factors influencing GPs' diagnostic decision-making and then examine the strength and stability of ratings of the factors' desirability and importance.

RESULTS: Thirty-nine factors were identified. Factors demonstrating high stability but no consensus included the importance of evidence-based medicine, the potential ramifications of a diagnosis, and the desirability of medicolegal issues. Factors for which there was disagreement in the first Delphi round but consensus in the second round included the importance of patient advocacy/support groups and the desirability of examination findings. Rasch analysis indicated that the questionnaire was close to the model (88.6% and 86.2% of variance in the ratings of importance and desirability explained).

DISCUSSION: Participants readily identified factors influencing GPs' diagnostic decision-making. Their ratings did not appear to support a prescriptive model of medicine, yet two cornerstones of prescriptive medicine, clinical information and probability of disease, were rated as highly desirable and important.

KEYWORDS: Decision-making; diagnosis; bias; Rasch analysis; general practitioners

Introduction

Making an accurate diagnosis and selecting an appropriate treatment can have profound consequences for an Accident Compensation Corporation (ACC) claimant, determining when or if they are able to return to work. Most investigations of factors influencing diagnostic or clinical decision-making employ quantitative methods rather than qualitative methods. Qualitative techniques can capture a range of factors which may be overlooked by quantitative methods.[1] In one such qualitative study dermatologists were asked to identify non-clinical or diagnostic factors and to indicate how influential the factors were.[1] Three types of factors were identified: patient-centred (e.g. patient choice, place of residence, ethnicity and age), clinician-related factors (e.g. time constraints, relationship with colleagues and staff, and pressure from the pharmaceutical industry), and practice-related factors (e.g. public or private treatment, cost of treatment to NHS).

United States studies have found regional and practice variations in diagnostic practices (e.g. overuse of diagnostic tests) which were related to the intensity of hospital and physician services.[2] Diagnostic practices were less likely to be related to patient characteristics.

Hajjaj et al.[3] argued that understanding how non-clinical factors affect evidence-based practice is important in clinical consultation, but usually overlooked. They argued that a combination of clinical and non-clinical factors may contribute to

J PRIM HEALTH CARE
2012;4(3):223–230.

CORRESPONDENCE TO:
Kathleen Callaghan
Human Factors Group, Department of Surgery, Faculty of Medical and Health Sciences, The University of Auckland, PB 92019 Auckland, New Zealand
k.callaghan@auckland.ac.nz

a clinician's decision without their awareness (e.g. patient adherence to advice and taking medication may influence the choice of treatment). They also suggested that a major challenge to practising medicine is integrating evidenced-based medicine (EBM) with important non-clinical factors while maintaining good standards of care. For example, a United Kingdom report on fitness for work advocated adopting the biopsychosocial approach to judging an employee's capacity or fitness for work rather than focusing on their physical symptoms.[4] Others have argued that while the biopsychosocial model is an inadequate scientific tool, it is nevertheless useful for clinical and teaching purposes[5] or the foundation of a philosophy of medicine.[6]

The current study employed qualitative and quantitative techniques to examine factors influencing diagnostic decisions. Three methods were used to identify and evaluate factors affecting diagnostic decisions (especially within an ACC-related context): focus groups, the Delphi method, and Rasch analysis. Using mixed quantitative and qualitative methods combines the strengths of the two while reducing their weaknesses.[7] The general practitioner's (GP's) evaluation of a patient's injury or injuries is critical to the success of an ACC claim: 'ACC legislation is a case in point where diagnosis routinely affects entitlement and cover decisions.'[8] A misdiagnosis can result in temporary or long-term loss of employment, reduced employment opportunities, and poorer health.[9,10] Analyses of ACC claims have revealed shortcomings in the diagnoses[11–13] so a greater understanding of decision-making is vital. Given the predominance of the biomedical model of medicine and the prescriptive and normative approaches to decision-making,[14,15] it was expected that the factors identified in the current study would reflect these models and approaches.

Method

Focus groups

Focus groups are 'a group of individuals selected and assembled by researchers to discuss and comment on, from personal experience, the topic that is the subject of the research.'[16] Focus groups may be biased if one individual dominates the discussion or if the participants are too readily influenced by the facilitator;[17] however, some authors believe such techniques are necessary to reveal hidden values and beliefs.[18]

The Delphi method

The Delphi method was used to identify and prioritise factors that influence GPs' decision-making. 'Delphi may be characterised as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem.'[19] 'The Delphi survey is a group facilitation technique, which is an iterative multi-stage process, designed to transform opinion into group consensus.'[20] The Delphi method reduces the influence of more dominant participants and group pressure.[21]

Rasch analysis

The Rasch model postulates that '...the probability of endorsing a statement is a logistic function of the difference between two independent quantities: the strength of the respondent's attitude and the location of the statement on the attitude continuum.'[22] Rasch analysis yields a measure on a unidimensional scale of the strength of a respondent's attitude, and another measure of the location on the same scale of the opinion expressed by an item under study.[23] The units of the scale are log-odds ratio or logits. In this study, the key respondent attitude is 'a tendency to endorse' the item and the key item measure is the importance or desirability of a diagnostic factor. The scale was calibrated such that the average of the items' importance or desirability was arbitrarily set to zero.

The University of Auckland Human Subjects Ethics Committee informed the investigator that ethical approval was not required.

Participants

Expert groups responsible for setting GP standards in NZ were identified. There are a number of groups that influence GP standards; however, many are biased, either legislatively or otherwise, towards a particular viewpoint. Academic organisations were identified as being the least biased 'standard setters'; therefore, participants were

selected from five academic organisations which provide GP training. These were the three Departments of General Practice at the University of Otago (Dunedin, Christchurch and Otago), the Departments of General Practice at The University of Auckland, and The Royal New Zealand College of General Practitioners (RNZCGP).

Heads of department were asked to select a minimum of four participants from senior members of their department currently in clinical practice. A reasonable gender balance and a mix of rural/city practice familiarity was preferred.

The characteristics of the 14 focus group participants were as follows: 10 were male, 4 were female; 13 were in current clinical practice (one had recently ceased practice to complete a full-time postgraduate degree); 11 were in urban practice, and 3 were in rural practice.

Focus group meetings

The investigator met with three of the five expert groups (time constraints prevented a meeting with the RNZCGP or the Christchurch Department of General Practice). Background information about the study was sent prior to the meetings. The investigator explained the purpose of these meetings, but did not participate in the subsequent discussion. The focus group from the Dunedin Department of General Practice comprised six participants while the other groups comprised four participants.

The lists of factors obtained from each meeting was examined for overlapping concepts and repetitions, and a composite list embracing the intent of all the expert groups consulted was compiled. An electronic questionnaire for Round 1 of the Delphi process was constructed from the composite list. Copies of the questionnaire together with instructions and background information were sent by email to 23 experts who had not participated in the focus groups. The questionnaire contained 39 factors identified by the focus groups as influencing GP diagnostic decision-making. The experts were asked to rate the importance and desirability of each factor using a seven-point scale, where 1 represented 'not at all important' or 'not at all desirable' and 7 represented 'very important', or 'very desirable'. Experts were free to comment on each factor.

WHAT GAP THIS FILLS

What we already know: Clinical and non-clinical factors may bias clinical decision-making. Bias is related to the GP's perception of the costs and benefits of making one choice over another.

What this study adds: Thirty-nine factors that potentially bias diagnostic decision-making by GPs and the subjective value placed on these factors were identified. Despite responses diverging from the prescriptive model, standard setters endorsed two cornerstones of the model—clinical information and probability of disease were rated as both highly important and desirable.

Despite both email and telephone reminders, only 12 Round 1 responses were received—a response rate of 52%. Of the 12 respondents, 6 were male, 6 were female; 11 were in current clinical practice, 1 had recently ceased practice to complete a full time postgraduate degree; 5 were in urban practice, and 3 were in rural practice.

Following standard Delphi methodology, the experts who responded in Round 1 were asked to re-rate each influencing factor and provide comments if they wished (Delphi Round 2). Several areas were clarified based on comments from the previous round, and one new factor added—a question designed to explore the influence of potential ramifications of the diagnosis.

Clarifications and alterations were clearly identified by a different coloured font. Experts were provided with their original ratings from Round 1 together with the group mean rating for each item. Only 11 Round 2 responses were received. Given the effort and time required to elicit completed Round 2 responses, it was assumed that the cost of a further Delphi round was likely to outweigh any benefits obtained; hence, a third round was not undertaken.

A Rasch analysis was performed using WINSTEPS® Version 3.55.[24] The method adopted was the rating-scale version of the Rasch model[25] where the ratings given to each item are not assumed to be equally spaced, but all items share the same structure.

Results

Thirty-nine factors were identified by the focus groups as influencing GPs' diagnostic decision-making (Tables 1 and 2).

Standard Delphi methodology

The summed mean ratings and stability or consensus of response for the group of GP standard setters are listed in Tables 1 and 2. Stability of response is an important, albeit often unreported, consideration in the analysis of Delphi responses. Whether or not further Delphi rounds may be productive can be established by measuring the stability of respondents' opinion distribution curves over successive rounds. The assessment of stable non-consensual distributions should be of equal interest in assessing opinion as the assessment of stable consensual distributions.

Group stability of opinion between the two Delphi rounds was assessed[26] and is reported in Tables 1 and 2. The absolute differences in the histograms of responses were calculated (subtracted column-wise) for two successive rounds and then summed to show total units of change. Net person changes were then calculated by dividing the total units of change by two (any one participant's change of opinion is reflected in the histogram differences by two units of change). Finally, the percentage change was calculated by dividing the net person changes by the number of participants. In this method, up to a 15% change level is accepted as representing a stage of equilibrium. The higher the percentage change, the less stable the group's position.

Responses from the participant who withdrew between Rounds 1 and 2 were eliminated from the analysis of stability. Similarly, factors which were not rated by one of the remaining participants could not be analysed accurately (no factor had more than one non-response). Stability was therefore represented as a range, with the lowest number representing the respondent as having maintained their previous rating and the highest number representing a change of rating.

Of interest are factors demonstrating high group stability without consensus (indicating fixed disagreement). The importance of EBM, potential ramifications of the diagnosis and the desirability of medicolegal issues are examples of these. In most Delphi analyses, consensus is assumed to have been reached when a specified percentage of responses falls within a prescribed range. Neither the proportion nor the range has been defined in the literature.[20] For this study, consensus was defined as having been reached when all responses fell within a three-point range at Round 2. If all responses fell within a two-point range (or less) then this was defined as strong consensus. Responses falling outside of a three-point range constituted disagreement. This is consistent with other interpretations of agreement and disagreement.[27]

Some factors showing consensus in Round 2 were highly unstable indicating that the group had shifted from a state of disagreement to a state of agreement between Round 1 and Round 2. Examples of this are the importance of patient advocacy/support groups and the desirability of examination findings as influencing factors.

It is often assumed that each item measured contributes in a meaningful way to the construct being investigated; that is, the questionnaire measures a single dimension. The dimensionality of the questionnaires used in this study was determined by principal components analysis. The Rasch measure explained 88.6% of the variance in ratings of importance (unexplained variance measured by the next component was 2.8%) and 86.2% of the variance in the ratings of desirability (unexplained variance measured by the next component was 4%). These results indicate that the questionnaire was unidimensional and close to the Rasch model's intent of measuring 'one attribute at a time'.[23] While the investigator labelled the constructs 'importance' and 'desirability', it can be inferred from examination of the items that the one attribute measured is represented by the label.

The assumption that ratings are equally spaced was tested using Masters and Wright's method[28] in which cumulative answers give the probability of responding greater or equal to each one of the available categories (e.g. ≥1, ≥2 etc). The threshold value is set where the cumulative probabilities equal 0.5. For example, in Table 3 the logit

Table 1. Summed mean ratings of importance, stability, degree of stability, consensus and strength of consensus

Influencing factor	Mean rating (importance)	Stability	Degree of stability	Consensus	Strength of consensus
History	6.8	Yes	9.1	Yes	†
Examination findings	6.2	Yes	9.1	Yes	*
GP's personal clinical experience	5.5	Yes	9.1	Yes	*
GP's knowledge of local conditions	5.4		18.2	Yes	†
Results of investigations	5.3	Possibly yes	13.6–18.2		
Evidence-based medicine	5.2	Yes	9.1		
Characteristics of the GP	4.8		22.7		
Need to achieve an outcome	4.4		31.8–36.4		
Patient expectations	4.2		36.4		
Patient advocacy	4.0		22.7–27.3		
Medicolegal issues	4.0		36.4		
Characteristics of the patient	4.0		27.3		
Reasonable patient pressure	4.0		27.3		
Closeness of GP/patient relationship	4.0	Possibly yes	13.6–18.2		
Time available for the consultation	4.0		18.2		
External feedback from a medical source	3.9		18.2		
Potential ramifications of the diagnosis	3.9	Yes	0.0		
The clinical setting	3.6		27.3		
Need to justify a course of action	3.6		18.2		
Personal circumstances of the patient	3.5		36.4		
Expectations of external medical professionals	3.5		22.7		
The Health and Disability Commissioner	3.1		18.2		
Potential implications for the wider community	3.1		50–54.5		
Context in which the diagnosis is made	3.1		18.2		
Administrative requirements	3.0		27.3		
Unreasonable patient pressure	3.0		36.4		
GP's desire to please the patient	3.0		27.3		
Diagnostic algorithms/categories/protocols	3.0		18.2		
Fear of uncertainty	2.9		36.4		
Whether or not treatment is available for the diagnosed condition	2.9		27.3		
GP's perception of what other external health professionals may think	2.8		36.4		
GP's perception of the state of the national health care system	2.6		27.3		
Marketing/media	2.6		18.2		
Legal requirements	2.6		27.3		
Patient advocacy/support groups	2.5		36.4	Yes	*
Technological tools	2.3		27.3		
Who is funding the consultation	1.7	Yes	9.1	Yes	*
Business considerations	1.6		18.2	Yes	†
External incentives	1.1	Yes	9.1	Yes	†

* Consensus
† Strong consensus

Table 2. Summed mean ratings of desirability, stability, degree of stability, consensus and strength of consensus.

Influencing factor	Mean rating (desirability)	Stability	Degree of stability	Consensus	Strength of consensus
History	6.6	Yes	0.0	Yes	†
Examination findings	6.2		27.3	Yes	*
Evidence-based medicine	6.2	Yes	9.1	Yes	*
GP's knowledge of local conditions	5.7	Yes	9.1	Yes	*
Results of investigations	5.5		18.2		
GP's personal clinical experience	5.2		27.3		
External feedback from a medical source	4.3		36.4		
Technological tools	3.8		31.8–36.4	Yes	*
Potential implications for the wider community	3.6		27.3		
Patient advocacy	3.6	Possibly yes	13.6–18.2		
Characteristics of the GP	3.5		18.2		
Characteristics of the patient	3.5		45.5		
Patient expectations	3.4		18.2		
Need to achieve an outcome	3.3		22.7–27.3		
Closeness of GP/patient relationship	3.1		40.9–45.5		
Medicolegal issues	2.8	Yes	9.1		
Expectations of external medical professionals	2.8		36.4		
Diagnostic algorithms/categories/protocols	2.8		27.3		
Legal requirements	2.6		45.5		
Need to justify a course of action	2.4		31.8–36.4		
Reasonable patient pressure	2.4		36.4		
Personal circumstances of the patient	2.3		22.7–27.3		
The Health and Disability Commissioner	2.3		27.3		
Potential ramifications of the diagnosis	2.3		18.2		
GP's perception of what other external health professionals may think	2.1		18.2		
Administrative requirements	1.9		18.2	Yes	*
The clinical setting	1.9		18.2		
Time available for the consultation	1.9		18.2		
Patient advocacy/support groups	1.9		36.4–40.9	Yes	*
Marketing/media	1.6	Possibly yes	13.6–18.2	Yes	†
GP's desire to please the patient	1.6		36.4	Yes	*
Fear of uncertainty	1.6		22.7–27.3	Yes	†
Whether or not treatment is available for the diagnosed condition	1.6	Possibly yes	13.6–18.2	Yes	*
Unreasonable patient pressure	1.4		18.2	Yes	†
Who is funding the consultation	1.3		18.2	Yes	†
Context in which the diagnosis is made	1.3	Yes	4.5–9.1	Yes	*
GP's perception of the state of the national health care system	1.2		27.3	Yes	†
Business considerations	1.2	Yes	9.1	Yes	†
External incentives	1.0	Yes	9.1	Yes	†

* Consensus
† Strong consensus

Table 3. Fifty percent cumulative probabilities for ratings of importance and desirability

Likert category (desirability)	1	2	3	4	5	6	7
50% cumulative probability (logits)	...	-1.97	-0.96	-0.27	0.38	1.02	1.81
Likert category (importance)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
50% cumulative probability (logits)		-2.77	-1.70	-1.00	0.15	1.62	3.70

value of -2.77 is the threshold at which a Likert rating of 1 is equally probable as a rating of 2 or above. The logit value of 3.70 is the threshold at which a rating of below 6 is as equally probable as a rating of 7. The thresholds are not spread equidistantly and it is easier to move between categories when rating desirability than when rating the importance of influencing factors.

When the distribution of participants and factors was plotted, the most important influencing factor was history, followed by examination findings, and the GP's personal clinical experience. However, the difference between history and examination findings is 2.22 logits versus 1.43 logits between examination findings and GP's knowledge of local conditions; hence, history would appear to be a much more important influencing factor than would be expected from knowledge of its ordinal position alone. Similarly, external incentives are regarded as being much less important than business considerations, with a distance of 1.89 logits separating these factors.

While history is viewed as the most desirable influencing factor, it is not as desirable as it is important, according to the difference in ratings (-5.95 logits for importance versus 3.56 logits for desirability). This indicates that the experts preferred to de-emphasise history as an influencing factor while maintaining its ordinal position. EBM was the sixth most important influencing factor but the second most desirable factor, despite representing only a difference of 0.91 logits. At the other end of the importance and desirability ordinal scale, factors such as external incentives, business considerations, and who is funding the consultation show very little logit mismatch.

Both the importance and desirability questionnaires are marked by a very broad span of 10.3 and 7.97 logits respectively, indicating a quite marked 'black and white' view of the factors at either extreme. The item reliability index was used to test the replicability of item placements (influencing factors) assuming these same factors were to be given to another group of comparable standard setters. For ratings of importance the item reliability index was 0.94 (an estimated 94% of the observed variance in the results obtained can be regarded as true variance, with the remaining 6% attributable to measurement error). In other words, the results obtained could be expected to correlate 0.94 with the results from a parallel measure of the same underlying construct. For ratings of desirability, the item reliability index was 0.93.

ACC-specific ramifications were considered less important than generic ramifications (0.01 versus 0.18 logits). The ramifications of a diagnosis (whether generic or ACC-specific) were not considered to be a desirable influence on diagnostic decision-making (0.34 and 0.39 logits).

Discussion

The relatively low number of participants in this study and the low response rate of 52% raises some concerns over the reliability[29,30] of the results obtained; however, the item reliability index values ameliorate these concerns.

The current study used a qualitative approach to identify factors influencing GPs' decision-making together with quantitative techniques. The stability of responses was also assessed. An advantage of Rasch scaling is that it enables meaningful comparisons between different groups undertaking the same questionnaire.

It is noteworthy that the standard setters readily identified 39 factors that influence GPs' diagnostic decision-making. The factors identified are not unexpected because they reflect everyday influences such as what others think of us, what we think of others, legality, need to achieve an outcome

and so forth. That these factors can be identified supports the controversial notion that it is not possible to divorce the practice of medicine from the society in which it is practised.[31] Interestingly, while a prescriptive view of diagnostic decision-making (in which decision-making is improved through the use of clinical guidelines etc.[32]) appears invalidated by the responses obtained in this study, the factors on which a prescriptive view of medicine relies (clinical information and probability of disease) are rated as both highly important and highly desirable. This apparent dissonance might be explained by Elstein:[33] 'All models and theories are simplifications of reality. Models are not reality however, and theory is not practice.'

Further research is required to determine the prevalence and consequences of diagnostic errors and to obtain further insight from an in-depth 'systems' approach (including nature of the error and underlying causal/contributory factors). A diagnosis relating to an ACC claim can be critical to whether a patient returns to work. Remaining in or returning to work is recognised as providing a significant contribution to health and well-being.[34–36] A better understanding of diagnostic decision-making within an ACC-related context, therefore, has important financial and social consequences not only for the patient but also for their community.

ACKNOWLEDGEMENTS
We thank John Irwin and Pat Bullen for their invaluable comments on this manuscript.

FUNDING
The source of funding for the study was a University of Auckland Senior Health Research Scholarship.

COMPETING INTERESTS
None declared.

References

1. Hajjaj F, Salek M, Basra M, Finlay A. Nonclinical influences, beyond diagnosis and severity, on clinical decision making in dermatology: understanding the gap between guidelines and practice. Br J Dermatol. 2010;163(4):789–99.
2. Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. N Engl J Med. 2010;363(1):45–53.
3. Hajjaj F, Salek M, Basra M, Finlay A. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. J R Soc Med. 2010;103(5):178–87.
4. Bevan S, Passmore E, Mahdon M. Fit for work? Musculoskeletal disorders and labour market participation. London; 2007.
5. Kendler KS. The rise and fall of the biopsychosocial model: reconciling art and science in psychiatry. Am J Psychiatry. 2010;167(8):999–1000.
6. Lewis B. The biopsychosocial model and philosophic pragmatism: is George Engel a pragmatist? Philos Psychiatr Psychol. 2008;14(4):299–310.
7. Johnson R, Onwuegbuzie A. Mixed methods research: a research paradigm whose time has come. Educ Res. 2004;33(7):14–26.
8. Callaghan KSN. Factors that influence general practitioner diagnostic decision-making and a comparison with other stakeholders. University of Auckland; 2006.
9. The Royal Australasian College of Physicians. Compensable injuries and health outcomes. Sydney: The Royal Australasian College of Physicians; 2001.
10. The Royal Australasian College of Physicians AFOEM. Helping people return to work: using evidence for better outcomes. Sydney; 2010.
11. Gorman DF. Audit of the ACC Workwise Clinic—December 1997. Auckland: The University of Auckland; 1997.
12. Gorman DF. Audit of the ACC Workwise Clinic—September 1997. Auckland: The University of Auckland; 1997.
13. Gorman DF, Jarvie P, Robinson P. Occupational health practice in New Zealand. NZ Med J. 1999;112:79–82.
14. Shaban RZ. Theories of clinical judgment and decision-making: a review of the theoretical literature. J Emerg Primary Health Care. 2005;3(1–2).
15. Wilson HJ. The myth of objectivity: is medicine moving towards a social constructivist medical paradigm? Fam Pract. 2000;17(2):203–09.
16. Powell RA, Single HM. Focus groups. Int J Qual Health Care. 1996;8(5):499–504.
17. Cox SJ, Cheyne AJT. Assessing safety culture in offshore environments. Saf Sci. 2000;34(1–3):111–29.
18. DeJoy DM. Behavior change versus culture change: divergent approaches to managing workplace safety. Saf Sci. 2005;43(2):105–29.
19. Linstone H, Turoff M. Introduction. In: Linstone H, Turoff M, editors. The Delphi method. Techniques and applications. Massachusetts: Addison-Wesley Publishing Company; 1975.
20. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32(4):1008–15.
21. Normand S-LT, McNeil BJ, Peterson LE, Palmer RH. Eliciting expert opinion using the Delphi technique: identifying performance indicators for cardiovascular disease. Int J Qual Health Care. 1998;10(3):247–60.
22. Irwin RJ, Irwin KC. Attitude towards mathematics of 11-year-olds: analysis of a questionnaire and its relation to mathematical achievement. In: Waugh RF, editor. Frontiers in educational psychology. New York: Nova Science Publishers; 2004:107–16.
23. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. Mahwah, New Jersey: Lawrence Erlbaum Assoicates; 2007.
24. WINSTEPS: Multiple-choice, rating scale, and partial credit Rashch analysis [program]. Chicago: MESA Press; 2000.
25. Andrich D. A rating formulation for ordered response categories. Psychometrika. 1978;43:561–74.
26. Scheibe M, Skutsch M, Schofer J. Experiments in Delphi methodology. In: Linstone H, Turoff M, editors. The Delphi method. Techniques and applications. Reading, Massachusetts: Addison-Wesley Publishing Company; 1975:263–83.
27. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. Int J Tech Assessment Health Care. 1986;2 (Part 1):53–63.
28. Masters GN, Wright BD. The partial credit model. In: van der Linden WJ, Hambleton RK, editors. Handbook of modern item-response theory. New York: Springer; 1997:101–21.
29. Frisbie DA. Reliability of scores from teacher-made tests. Educ Meas: Issues and pract. 1988;7(1):25–35.
30. Traub RE, Rowley GL. Understanding reliability. Educ Meas: Issues and pract. 1991;10(1):37–45.
31. Gorman DF, Scott J. The social distortion of medical practice. Med Today. 2003;4(11):75–77.
32. Shaban RZ. Theories of clinical judgment and decision-making: a review of the theoretical literature. J Emerg Primary Health Care. 2005;3(1–2):23–33.
33. Elstein AS. Clinical problem solving and decision psychology: comment on 'The epistemology of clinical reasoning'. Acad Med. 2000;75(10):S134–S36.
34. Bevan S, Quadrello T, McGee R, Mahdon M, Vavrovsky A, Barham L. Fit for work? Musculoskeletal disorders in the European workforce. London; 2009.
35. Black DC. Working for a healthier tomorrow. Norwich; 2008.
36. The Royal Australasian College of Physicians AFOEM. Realising the health benefits of work. Sydney; 2010.