

The Dutch knowledge test for general practice: issues of validity

Yvonne D van Leeuwen, Marjan C Pollemans, Saskia SL Mol, Just AH Eekhof, Richard Grol, Maria J Drop

Objectives: This article describes a study performed to validate a nation-wide knowledge test for general practice. This test is administered to all GP-trainees in The Netherlands at fixed intervals during the time of their training. The test is a paper-and-pencil test with a progress testing format. Content and construct validity of the test are examined.

Methods: The content validity of the test is investigated by examining the procedure of test and item construction, and by defining a multidimensional blueprint. Construct validity is investigated by comparing mean scores of different groups: medical students, trainees at six different training levels and qualified general practitioners.

Results: A test blueprint is constructed which covers the domain of general practice care. The main dimension of the blueprint covers complaints and diseases, whereas additional dimensions provide an adequate age distribution of patient cases in the test, coverage of different aspects of the consultation, and the provision of enough cases dealing with emergency and chronic problems.

The mean group score increases with training level. The mean score of trainees at the end of training surpasses the mean score of qualified general practitioners.

Conclusions: It is concluded that the test assesses knowledge that is closely related to the GP's daily work. The test enables the monitoring of growth in

knowledge during postgraduate training. Content and construct validity of the test are satisfactory.

Eur J Gen Pract 1995; 1: 113-7.

Summary

The Knowledge Test for General Practice is a written test administered to all trainees in The Netherlands three times a year, which implies that it offers trainees the opportunity to detect their strengths and weaknesses on a regular basis and to compare these with those of their colleagues. The test consists of about 80 patient cases with a total of 160 items. A validity study has been performed with the emphasis on content validity, which is often erroneously assumed to be existent. The procedure of test construction was evaluated, as well as the format of cases and items, whereas a test blueprint was constructed. The test was assessed on relevance for general practice. The construct validity was examined by comparison of scores of groups of trainees at different training levels, GPs and medical students. The results warrant the conclusion that the test assesses knowledge that is closely related to the daily work of the general practitioner and that it has the potential to illustrate growth in knowledge during training.

Introduction

In the context of postgraduate training for general practice the assessment of the competence of trainees is of major importance. Training institutions have the obligation to ascertain that trainees have the potential to perform well as a result of training, thereby giving account of their efforts to protect the public against incompetence. Moreover, the competence of trainees reflects the quality of the training; assessment of competence may point out strengths and weaknesses of the training programme. Assessment instruments, therefore, should reflect the demands of the job and the learning objectives of education.

The Dutch institutes for postgraduate training for general practice have developed several tests to assess the trainees' competence: a technical skills test, a consultation skills test and a test for the assessment of knowledge relevant for daily practice. The quality of the latter has been investigated. Considerable attention was given to the issue of validity which in the case of written knowledge tests is often questioned.¹ In this paper the format of the test and the process of validation will be described.

Yvonne D van Leeuwen, general practitioner, research fellow.
Marjan C Pollemans, medical educationalist.
Richard Grol, professor of general practice.
Maria J Drop, professor of medical sociology.
University of Limburg, PO Box 616, 6200 MD Maastricht, The Netherlands.
Saskia SL Mol, general practitioner, research fellow.
University of Utrecht, PO Box 80045, 3508 TA Utrecht, The Netherlands.
Just AH Eekhof, general practitioner, research fellow.
University of Leiden, PO Box 2088, 2301 CB Leiden, The Netherlands.
Correspondence to: Dr Yvonne van Leeuwen.

Submitted: March 17th, 1995; accepted in revised form: August 9th, 1995.

Postgraduate training for general practice in The Netherlands

During postgraduate training for general practice in The Netherlands the trainee predominantly works and learns in general practice under the supervision of a GP trainer. The training period in hospital is restricted to eight months at most. Once a week trainees visit their training institution to attend a 'day release course'.²

The core element of the assessment of the trainees concerns their skills and attitudes 'on the job', which are evaluated by the trainer. The trainer's judgement is decisive for overall pass/fail decisions. Up until now, the knowledge and skills tests merely had a diagnostic and educative function.

The Knowledge Test for General Practice

The Knowledge Test for General Practice, introduced in 1987, is administered to all (about 500) trainees in general practice in The Netherlands at fixed intervals six times during their training. All trainees take the same test regardless of their training level. The test is set at the level of the qualified general practitioner at the moment of certification. The test is thus designed to record progress during training.³

The test consists of approximately 80 patient cases with a total of 160 case-related items.⁴ Figure 1 presents two examples of cases and their corresponding items.

The test material is constructed by staff members of the eight departments of general practice in The Netherlands and reviewed on relevance, content and syntax. After test administration the psychometric characteristics of the items, as well as the trainees' comments are examined. Bad psychometric item characteristics include for example a majority of incorrect answers on a specific item, or a high percentage of trainees failing this specific item, who otherwise score high on the test as a whole. An obvious explanation for these findings, to which the trainees' comments (such as contradictory literature references, ambiguity in the phrasing of the item) may give a clue, is conditional for elimination.

The feedback to the trainees consists of their own scores, the mean score of their own training group (12 trainees) as well as the mean score of all Dutch trainees of the same training level.

Aim of the study

The aim of the present study was to determine the validity of the knowledge test.⁵⁻⁸

Content validity was our main concern: are the problems included in the test similar to the problems trainees meet in practice before or after certification?⁹ Do items refer to other than trivial facts which are so often the 'main course' of written knowledge tests?¹

Construct validity, however, is likewise important:¹⁰ do the test results fit into our assumption of the 'acquisition of general practice knowledge'? In this case the assumption

Figure 1. Examples of cases and items.

Charles Evert, 16 years old, visits his GP with the following complaint: a severe pain in his right testicle. The pain began acutely, about an hour ago. Examination reveals a red and swollen scrotum. Palpation is painful. It is impossible to demarcate the epididymis from the testis. Lifting the testis increases the pain.

In this case an epididymitis is more likely than a torsion of the testis. (true/false/?)

Reference: Boer J de. Textbook of Urology. Utrecht: Bunge, 1989: 89.

Mrs Cleveland, 75 years old, is known to have an advanced arthrosis of her left knee. She enters the room, walking slowly and leaning to the right on her stick. The left knee is swollen and warm. There is an effusion.

Among the therapeutic measures, appropriate at this moment is/are:

- application of icepacks (true/false/?)
- massage and exercises (true/false/?)
- intra-articular injection of corticosteroids (true/false/?)
- holding the stick in the other hand (left) (true/false/?)

Reference: Wolff AN de. Arthrosis. *Gen Pract* 1988; 3: 65-9.

is, that the level of knowledge increases with experience in general practice.

The following research questions were formulated.

Does the test content reflect the knowledge that (future) general practitioners need to solve problems encountered in general practice (content validity)?

Do the test results reflect the assumed increase in knowledge during postgraduate training (construct validity)?

These questions may be operationalised as follows. The test should cover the domain of general practice care, consist of cases that a general practitioner encounters in daily practice, and focus on dilemmas that frequently arise in the context of these cases: the 'heart of the matter' or key features of the problem.¹¹

If the knowledge that is assessed is acquired during training, then test scores should increase with training level.

Methods

Methods advocated in the literature to ensure content validity of a test are: job description, expert judgement and the 'critical incident technique'.⁸

The last method implies that the test content should be composed on the basis of what is crucial in daily care: issues are selected which should be mastered if patients are not to run a considerable risk of dying or of suffering severe complications.¹²

We adopted the first two, not formulating the test only on the basis of critical incidents, regarding such an approach as conflicting with the idea of covering the domain.

Table 1. Blueprint main dimension (chapters and percentage of items for each chapter).

Blueprint main dimension	Number of items	%
General and unspecified (ICPC A)	8	5
Blood and blood-forming organs and lymphatics (spleen, bone marrow) (ICPC B)	2	1
Digestive (ICPC D)	10	6
Eye (ICPC F)	6	4
Ear (ICPC H)	6	4
Circulatory (ICPC K)	16	10
Musculoskeletal (ICPC L)	18	11
Neurological (ICPC N)	8	5
Psychological (ICPC P)	8	5
Respiratory (ICPC R)	16	10
Skin (ICPC S)	12	7
Endocrine, metabolic and nutritional (ICPC T)	6	4
Urological (ICPC U)	6	4
Pregnancy, child-bearing, family planning (ICPC W)	6	4
Female genital (including breast) (ICPC X)	6	4
Male genital (ICPC Y)	6	4
Theoretical issues (including ICPC Z: social)	20	12
Total	160	100

The regular test of February 1992 was applied in this study to establish construct validity. The correct-minus-incorrect scores of trainees of six successive training levels - at four month intervals - were compared. Medical students in their general practice clerkship and qualified general practitioners (mostly trainers) participated as reference groups. The test takers evaluated the test on relevance by completing a questionnaire with a five-point Likert scale.

Results

Content validity

1. Construction of a test blueprint

In order to cover the domain of general practice it is vital to define this domain. A job description for Dutch general practitioners, agreed upon by the profession, was available.^{13,2} However, a more detailed description of the intended domain was needed to serve as a basis for the blueprint. The literature provided only few examples.^{14,15}

General practice knowledge being a concept which can be defined according to different dimensions (e.g. morbidity, sorts of care, characteristics of the population), seemed to demand a test with a multidimensional blueprint.

The American Board of General Practice had adopted such a blueprint.¹⁵ The choice of chapters, however, did not meet our criteria on several points. Among others, the choice for the International Classification of Diseases as the main dimension seemed inappropriate. We adopted the International Classification of Primary Care (ICPC) instead,¹⁶ considering it a better representation of the domain of general practice and of the educational objectives of postgraduate training. To include non-clinical aspects a chapter 'theoretical issues' was added, covering subjects

like 'law and medicine', ethics, patient education, medical decision making and social problems.

The number of questions for each ICPC chapter was fixed by consensus, taking account of morbidity, the variety of problems within each chapter and the challenge that the different topics offer to the general practitioner. Additional dimensions were chosen on the basis of relevant topics in general practice, like care for the elderly and chronic diseases. Definitions for the chapters of these dimensions were formulated.

Table 2. Blueprint additional dimensions (chapters and percentages of cases or items for each chapter).

Blueprint additional dimensions	%
Age of patient	cases
under 15	10 - 25
15 - 75	50 - 80
above 75	10 - 25
items	
Aspects of consultation	items
Diagnosis	min 40
Treatment	min 40
medication	min 20
otherwise	min 20
Morbidity rate	max 10
Miscellaneous	max 10
Urgency of the problem	cases
Emergency	5 - 10
No emergency	90 - 95
Course of illness	cases
Chronic diseases	10 - 25
Otherwise	75 - 90

Table 3. Judgement of test takers concerning the content of the knowledge test of February 1992 (in percentages).

	Students (n=108)	Trainees (n=445)	GPs (n=365)
Relevance of subjects			
- (highly) relevant	92	85	90
- undecided	8	12	8
- (highly) irrelevant	0	3	2
Relevance of items			
- (highly) relevant	83	73	84
- undecided	14	18	11
- (highly) irrelevant	3	9	5

An expert group, members of the boards of the Dutch College of General Practitioners and official representatives of the eight training institutes for general practice (25 all together), evaluated the procedure and the suggested blueprint. This resulted in several alterations.¹⁷ The definite blueprint is presented in table 1 and 2.

2. Content validity of cases and items

The authors of test cases and items, all GPs, were instructed to describe cases they encountered in practice and to relate the test items to the 'key feature' of the problem.¹⁸

Every item was checked on relevance, content and syntax by an independent group of GPs. This procedure enhances content validity.¹⁹

The test was evaluated by the test takers, immediately after administration. This 'consumers opinion' is an additional

Table 4. Number of participants within each group (N), their corresponding mean scores (percentage correct minus incorrect) and standard deviation (SD).

Participants	N	mean score (% C - IC)	SD
Medical students	108	23	9
Trainees**	445		
level 1 (2 months in training)	85	34 *	10
level 2 (6 months in training)	103	39 *	11
level 3 (10 months in training)	46	40	10
level 4 (14 months in training)	55	40	9
level 5 (18 months in training)	99	46 *	11
level 6 (22 months in training)	57	49 *	10
GPs	365	45 *	11
Total number of participants	918		

* significant difference ($p < 0.05$) with preceding level.

** at the time of this study postgraduate training consisted of a two year training period. In September 1994 it was been extended to three years.

check on content validity. Table 3 illustrates the appreciation of the test takers for the subjects and items of the February 1992 test. This high approval rating is a consistent finding.

Construct validity

Table 4 presents the number of participants within each group as well as the corresponding correct- minus- incorrect scores and standard deviation.

Trainees at the start of training scored considerably better than medical students during their clerkship in general practice. The mean score increases with the duration of postgraduate training. The mean score of trainees at the end of training (level 6) surpasses that of qualified general practitioners.

Discussion

Test validation is essential to ensure that candidates are assessed on what is relevant for their (future) professional performance. All too often tests are applied without scrutiny of content and format or taken at face validity. The present study shows an example of a stepwise and soundly based validation procedure.

The conclusion is warranted that the test content reflects general practice, which allows inferences of test scores in terms of level of 'general practice knowledge'. The fact that it is a written test with a closed answer format implies that recognition and not recall of information is assessed. This 'disadvantage' is, however, an acceptable 'price' for nationwide assessment of large groups of candidates and does, according to the literature, not influence validity substantially.²⁰

The empirical data confirm the hypothesis that the test discriminates between training levels and has the potential to illustrate growth. The 'plateau' at training level 3-4 needs further explanation. Many trainees spend their time in hospital in this period, which restricts their opportunity to study morbidity as presented in general practice.

It is noteworthy that the scores of trainees at the end of training surpass the scores of general practitioners. This finding has repeatedly been confirmed. It may be that the trainee is favoured over the practising GP by the training setting: the trainee does not learn exclusively from seeing patients but also from discussions with the trainer and with experts and peers during the weekly day release courses. Discussion and reflection may contribute to integration and consolidation of recently acquired knowledge. Consequently, GPs may rely more on patient-related knowledge instead of on problem-related knowledge.

It is reassuring that practising GPs appreciate the test as relevant to general practice and as useful for assessing their own knowledge. These statements reflect a need for knowledge of the 'state of the art', as long as its relation to daily practice is obvious. ■

References

- 1 Marshall J. Assessment during postgraduate training. *Acad Med* 1993; 68: 23-6.
- 2 Dubois V, Everwijn S, Geldorp G van, et al. The construction of a new curriculum of vocational training for general practice in The Netherlands. Utrecht: Royal Dutch Medical Association, 1987.
- 3 Verwijnen GM, Imbos Tj, Snellen H, et al. The evaluation system of the medical school of Maastricht. *Assess Evaluat Higher Ed* 1982; 3: 225-44.
- 4 Leeuwen YD van, Hessen PAW van. Clinical competence and objective questions: Tactics to realize a true/false format assessing competence. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, editors. *Teaching and Assessing Clinical Competence*. Groningen: BoekWerk publications, 1990.
- 5 Berkel HJM van. De diagnose van toetsvragen [dissertation]. Amsterdam: Centrum voor Onderzoek van het Wetenschappelijk Onderwijs (COWO), Universiteit van Amsterdam, 1984.
- 6 Neufeld VR, Norman GR, editors. *Assessing Clinical Competence*. New York: Springer Publishing Company, 1985.
- 7 Carmines EG, Zeller RA. *Reliability and validity assessment*. Beverly Hills/London: SAGE publications, 1985.
- 8 Kane MT. The validity of licensure examinations. *Am Psychol* 1982; 8: 911-8.
- 9 Ebel R. The practical validation of tests of ability. *Educ Meas, issues pract* 1983; 2: 7-10.
- 10 Cronbach LJ. What price simplicity? Comment on: The practical validation of tests of ability [Ebel R. *Educ Meas, issues pract* 1983; 2: 7-10]. *Educ Meas, issues pract* 1983; 2: 11-2.
- 11 Bordage G, Page G. An alternative approach to PMP's: the 'key features' concept. In: Hart IR, Harden RM, editors. *Further Developments in Assessing Clinical Competence*. Montreal: Heal Publications, 1987.
- 12 Flanagan JC. The critical incident technique. *Psychol Bull* 1954; 1: 327-57.
- 13 Springer MP, editor. *Basistakenpakket voor de huisarts* [Basic job description for the general practitioner]. Utrecht: Landelijke Huisartsen Vereniging, 1983.
- 14 Lockie C, editor. *Examination for Membership of the Royal College of General Practitioners (MRCGP)*. Development, current state and future trends [occasional paper 46]. London: Royal College of General Practitioners, 1990.
- 15 Pisacano NJ, Veloski JJ, Brucker PC, Gonnella JS. Classifying the content of board certification examinations. *Acad Med* 1989; ii: 149-54.
- 16 Lamberts H, Wood M, editors. *ICPC. International Classification of Primary Care*. Oxford: Oxford University Press, 1987.
- 17 Pollemans MC, Leeuwen YD van, Düsman H, Eekhof JAH, Mol SSL. *Achtergronden blauwdruk huisartsgeneeskundige kennistoets* [Background of the blueprint of the Dutch knowledge test for general practice]. Utrecht: Samenwerkingsverband IOH, 1993.
- 18 Kramer AWM, Pollemans MC, Leeuwen YD van. Het toetsen van kennis in de beroepsopleiding tot huisarts [Assessment of knowledge in vocational training]. *Med Contact* 1990; 7: 220-2.
- 19 Swanson D, Case S. Trends in written assessment: a strangely biased perspective. In: Harden R, Hart R, Mulholland H, editors. *Approaches to the Assessment of Clinical Competence*. Dundee: Centre for Medical Education, 1992.
- 20 Vleuten C van der, Newble D, editors. *Methods of assessment in certification*. In: Newble D, Jolly B, Wakeford R, editors. *The Certification and Recertification of Doctors: Issues in the Assessment of Competence*. Cambridge: Cambridge University Press, 1994.

ABSTRACT

Effect of maternal cigarette smoking on pregnancy complications and sudden infant death syndrome

Joseph R DiFranza, Robert A Lew

Background. The purpose of this study was to estimate the annual morbidity and mortality among fetuses and infants that can be attributed to the use of tobacco products by pregnant women.

Methods. Published research reports identified by literature review were combined in a series of meta-analyses to compute pooled risk ratios, which, in turn, were used to determine the population-attributable risk.

Results. Each year, use of tobacco products is responsible for an estimated 19,000 to 141,000 tobacco-induced abortions, and 32,000 to 61,000 infants who require admission to neonatal intensive care units. Tobacco use is also annually responsible for an estimated 1,900 to 4,800 infants deaths resulting from perinatal disorders, and 1,200

to 2,200 deaths from sudden infant death syndrome (SIDS).

Conclusion. Tobacco use is an important preventable cause of abortions, low birthweight and deaths from perinatal disorders and SIDS. All pregnant women should be advised that smoking places their unborn children in danger. The low success rate of smoking cessation among pregnant women suggests that efforts to reduce the complications of pregnancy attributable to tobacco use by pregnant women should focus on preventing nicotine addiction among teenage girls. ■

J Fam Pract 1995; 40: 385-94.