Semi-structured interview is a reliable and feasible tool for selection of doctors for general practice specialist training

Jesper Hesselbjerg Isaksen[1], Niels Thomas Hertel[2] & Niels Kristian Kjær

ABSTRACT

INTRODUCTION: In order to optimise the selection process for admission to specialist training in family medicine, we developed a new design for structured applications and selection interviews. The design contains semi-structured interviews, which combine individualised elements from the applications with standardised behaviour-based questions. This paper describes the design of the tool, and offers reflections concerning its acceptability, reliability and feasibility.

MATERIAL AND METHODS: We used a combined quantitative and qualitative evaluation method. Ratings obtained by the applicants in two selection rounds were analysed for reliability and generalisability using the GENOVA programme. Applicants and assessors were randomly selected for individual semi-structured in-depth interviews. The qualitative data were analysed in accordance with the grounded theory method.

RESULTS: Quantitative analysis yielded a high Cronbach's alpha of 0.97 for the first round and 0.90 for the second round, and a G coefficient of the first round of 0.74 and of the second round of 0.40. Qualitative analysis demonstrated high acceptability and fairness and it improved the assessors' judgment. Applicants reported concerns about loss of personality and some anxiety. The applicants' ability to reflect on their competences was important.

CONCLUSION: The developed selection tool demonstrated an acceptable level of reliability, but only moderate generalisability. The users found that the tool provided a high degree of acceptability; it is a feasible and useful tool for selection of doctors for specialist training if combined with work-based assessment. Studies on the benefits and drawbacks of this tool compared with other selection models are relevant.

FUNDING: not relevant.

TRIAL REGISTRATION: not relevant.

Selecting the right doctors for specialist training in family medicine is challenging. In the UK, there has been growing concern over high postgraduate drop-out rates [1], and a model involving test centres has been developed [2]. Studies show that these test centres exhibit a higher reliability and validity than interviews do [3-5]. But the model is resource demanding [2]. Other less manpower demanding selection tools are also in use, e.g. the Multiple Mini Interview (MMI). The MMI appears to be a feasible method, and it was recently introduced as a selection tool in several specialities [6].

In Denmark, a 6-month introduction position with work-based assessment (WBA) combined with unstructured interviews has been used for selection of future general practitioners (GPs) [7]. Due to the risk of low inter-rater reliability in the WBA, the selection process was optimised with a structured application form and a structured interview.

An employment interview may be a more valid indicator of future performance than previously assumed [8], and structured interviews seem to have a higher validity than unstructured interviews [9]. The more strict structure in the interviews may, however, elicit negative reactions from the applicant [10]. Little is known about the value of combined WBA and structured interviews.

We developed a new design for structured applications and interviews, and tested their applicability and feasibility as seen from both the applicants' and the assessors' perspectives. This paper describes the design of this selection tool and provides reflections concerning its acceptability, reliability and feasibility.

MATERIAL AND METHODS

In the Danish specialist training programme, the trainee has to fulfil criteria within seven areas of competence before specialisation [7]. We therefore decided to develop a selection model based on these seven key roles.

The interview model was developed during three application rounds for educational positions in specialist training in family medicine during the 2006-2007 period on the island of Funen, Denmark (**Figure 1**).

In the third application round, we used the final version of the interview guide based on the results from interviews and the experiences of the panellists/interview panel obtained during the first two rounds.

The interviews contained seven sections each related to one of the key roles. Each section had individual questions based on the written application, and standardised behavioural questions which were used for all applicants.

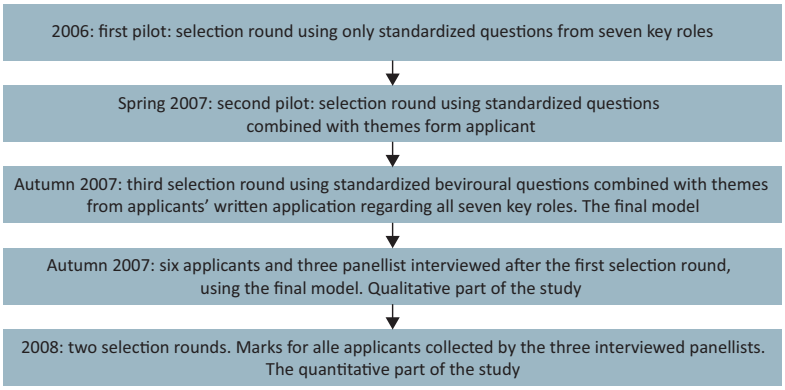
ORIGINAL ARTICLE

1) The Postgraduate Deanery and Office of Medical Education, Faculty of Medicine, University of Southern Denmark
2) H.C. Andersen Children's Hospital, Odense University Hospital
3) Unit of General Practice, Institute of Public Health, University of Southern Denmark

Dan Med J 2013;60(9):A4692

FIGURE 1

Study flow diagram.



During the interview, the members of the selection panel independently rated the applicant on a five-point rating scale within each role. The final result for the applicant was the average score of all ratings.

Testing the tool

After the final round, six applicants were randomly selected for individual in-depth interviews. Three of the applicants had been accepted for the training programme and three had been rejected.

The three panellists, two GPs and one junior doctor, were interviewed individually. The interviews followed a semi-structured interview guide, focusing on acceptability, perceived fairness, user friendliness and usability of the selection procedure. All interviews were recorded and transcribed.

An empirical thematic analysis with a grounded theory approach was used to analyse the interviews [11]. The statements were divided into meaning-carrying units and then independently categorised by two researchers, one who had and one who had not participated in the development of the selection procedure. Their pre-understanding and expectations were clarified before data analysis. Results were only included in case of agreement between both researchers in the interpretation of the data. The process allowed for researcher triangulation [12].

We combined qualitative data from the accepted and rejected applicants and data from the panellists in order to achieve data triangulation [12]. The results were furthermore put into perspective by including findings from the literature.

In order to make a comprehensive presentation of the qualitative results, data were condensed into themes.

The ratings from two selection rounds in 2008 in which the final model was used were analysed in IBM SPSS Statistics ver. 19. Reliability analysis was performed by calculating Cronbach's alpha, which is a measure of the internal consistency of a test score. Cronbach's alpha is considered acceptable if the value is above 0.70. Cronbach's alpha was calculated for each interviewer and for each selection round.

In order to assess the generalisability of the selection model, the G coefficient was calculated by using the GENOVA programme [13]. Generalisability refers to the extent to which the results of a study apply to individuals and circumstances beyond those studied. A G coefficient of 0.8 or above is considered excellent [13]. The results from a G study can also be used to estimate the number of questions needed to obtain an acceptable G coefficient.

Trial registration: not relevant.

RESULTS

Qualitative results

During the analysis three themes emerged: Acceptability, importance of the structure and fairness.

Acceptability

Both applicants and panellists generally accepted the interview as a tool for selection. The applicants found it natural and fully acceptable that the interview started with questions addressing their own written application.

The applicants felt that the overall interview duration was appropriate, although some requested more time for specific issues.

The applicants stated that acceptability would increase if they had been given an opportunity to make a supplementary comment at the end of the interview. The assessors, however, stated that these additional comments could not be part of the rating because it was not a compulsory part of the interview.

... I think that you show who you are much better [in an interview] than when you are sitting with the text at home. You can write a lot on the item when you have a long time to think about it.

I think you give a more accurate picture of yourself.

(Applicant interview 2)

Importance of the structure

The applicants agreed that discussing elements from their own application form and combining these with standardised behaviour-based questions gave fair insight into their qualifications. This combination made the method appear fair, and the interview method was widely described as pleasant.

The selectors found that this interview model with a combination of individualised and standardised ques-

! TEXT BOX

Content of the application for specialist training

Applicant's curriculum

Referees' report from basic training programme

Personal statement of reasons for application

tions worked well, although the method was more time-consuming in the preparation of the interviews. The planning of questions had the advantage that the interviews became more specific and directed more at the particular applicant's qualifications.

The applicants reported that the standardised questions could give rise to some anxiety. Questions related to clinical experience could be difficult to answer, especially for applicants who had only little experience in general practice yet. The applicants, therefore, stressed the importance of a high level of information before the interview, but overall they preferred being assessed in an interview compared to a selection process based on a written application only.

The assessors found that the structured interviews were much more effective than the previously used non-structured interviews. They only experienced minor shortcomings with the new structured interviews. They expressed concern that increased structure may result in the loss of important, unexpected details and in some situations might reduce the applicants' opportunity to form their own independent impression. The benefits of the structure, nonetheless, clearly outweigh this disadvantages. The interview guide, however, should not be too formal, particularly in interviews with poor performers, the interview guide should allow for further investigation/discussion.

Furthermore, the panellists had an impression that the method provided an acceptable reliability with regard to scoring/rating, especially for applicants with high or poor trainability.

… the method made the selection interviews more intriguing and it gave us a more profound insight ... and more nuanced picture of the applicant than we got in the old selection interviews. (Assessor GP)

… I think it [the interview] covers 80% of relevant issues. I think that maybe there are some personal motives that you don't tell. (Applicant interview 1)

Fairness

Both applicants and assessors found the rating of the ability to reflect very valuable. Confidence in the method was generally high.

... I think it is fair because you have to present your reflections instead of just saying: Yes, I have reflected. It becomes more specific; like that it becomes more convincing, also for the interviewer.

(Applicant interview 3)

... But it is also quite ok to be challenged [by the question]... I must admit I was a bit surprised the first time I was asked, "how do you react in a stressed situation?" … But I actually feel it was OK.

(Applicant interview 2)

The applicants believed that all key roles should be given the same weight in the selection process in order to secure diversity among future GPs. The panellists had the impression that the objectivity in the rating was superior to that achieved when using the previous interviews and that the method clearly provided a better opportunity to specify their impressions of the applicant.

Compared to the old method, the new method appears more fair and better ensures that personal matters do not influence the selection process.

(Assessor, trainee representative)

The assessors found that the seven key roles were a suitable framework for selection; however, they had concerns about the ratings of the roles "medical expert" and partly "communicator".

They stated that rating of clinical performance need structured clinical observation and that it could not be based only on an interview or on self-reported achievements. Concern was also expressed as to whether the communicative skills demonstrated in an interview would be representative of the applicants' communicative skills in clinical settings.

Quantitative results

Reliability

Marks given during the two rounds can be seen in **Table 1** and **Table 2**. We found a high Cronbach's alpha in the two sessions (**Table 3**). In the first interview session, we

! FIVE-POINT RATING SCALE

1. Poor and insufficient understanding and insight into the role
2. Hardly acceptable with one or several serious needs or concerns for insight into the role
3. Acceptable with several minor needs or concerns concerning insights into the role
4. Good understanding with a few minor or insignificant needs for insight into the role
5. Excellent

TABLE 1

Marks given by the two selections.

	Assessor A						Assessor B						Assessor C					
	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
2008-1																		
Applicant no. 1	3	3	4	3	4	3	4	4	4	3	4	4	4	4	4	4	4	4
Applicant no. 2	4	4	4	3	4	4	3	3	3	3	3	3	4	4	4	3	4	3
Applicant no. 3	3	3	4	4	3	3	3	3	2	3	3	3	4	3	3	4	3	4
Applicant no. 4	3	4	3	3	3	3	4	4	3	4	3	3	3	4	2	4	4	3
Applicant no. 5	4	4	4	4	4	4	4	4	4	4	5	4	4	4	5	5	5	5
Applicant no. 6	3	4	4	4	4	2	3	3	3	3	3	2	4	4	4	4	4	3
Applicant no. 7	2	2	3	2	2	2	2	3	2	2	2	2	2	3	3	3	3	3
Applicant no. 8	4	4	5	5	4	5	4	5	4	5	5	5	4	5	5	5	5	5
Applicant no. 9	2	2	3	3	2	2	2	2	2	2	2	2	3	2	3	3	2	3
Applicant no. 10	4	5	5	5	4	5	4	3	4	4	4	4	3	4	4	4	4	5
Applicant no. 11	3	3	3	3	3	3	3	3	4	4	3	3	4	3	3	3	3	3
Applicant no. 12	3	4	3	3	4	3	3	3	3	3	3	3	3	4	4	4	4	4
Applicant no. 13	4	3	4	4	3	3	3	3	3	3	3	3	4	3	4	4	4	3
2008-2																		
Applicant no. 14	3	3	3	3	3	3	4	4	5	4	4	3	4	4	4	4	5	4
Applicant no. 15	4	4	3	3	4	4	4	3	4	4	4	4	4	4	4	4	5	5
Applicant no. 16	3	3	3	4	5	4	4	3	3	3	3	3	3	3	4	5	4	4
Applicant no. 17	3	3	3	4	3	3	3	3	4	4	3	4	4	4	3	4	4	5
Applicant no. 18	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Applicant no. 19	4	4	5	4	4	4	3	4	5	4	4	4	4	3	3	4	4	4
Applicant no. 20	4	3	4	4	4	4	3	4	4	4	3	3	4	3	3	4	4	4
Applicant no. 21	3	3	3	4	4	5	3	3	3	3	3	3	4	3	3	3	4	3
Applicant no. 22	3	4	4	4	4	4	3	4	4	4	3	3	3	3	4	4	4	3
Applicant no. 23	3	3	3	3	3	3	3	2	3	3	3	3	3	2	3	3	3	3
Applicant no. 24	4	3	3	3	3	3	3	3	4	3	4	3	3	3	3	3	4	3
Applicant no. 25	3	3	4	4	4	3	3	3	3	3	3	3	3	3	4	4	4	3
Applicant no. 26	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	4	3	3
Applicant no. 27	2	4	4	4	3	3	4	5	4	5	4	4	4	5	4	5	4	5
Applicant no. 28	3	3	3	3	3	3	3	3	3	4	4	3	3	4	3	3	4	3
Applicant no. 29	3	5	4	4	4	4	4	4	4	4	5	5	3	5	4	4	4	4

A = medical expert; B = academic; C = professional; D = communicator; E = leader and collaborator; F = health promoter.

TABLE 2

Mean role values.

	Mean value	Median	Min.	Max.
Medical expert	3.499	3	2	5
Academic	3.532	3	2	5
Professional	3.546	3	2	5
Communicator	3.552	3	2	5
Leader and collaborator	3.533	3	2	5
Health promoter	3.506	3	2	5

TABLE 3

Cronbach's alpha calculations.

	Assessor A	Assessor B	Assessor C
Round 2008-1	0.931	0.957	0.890
Round 2008-2	0.762	0.850	0.835

found an acceptable G coefficient of 0.736 using generalisability theory, but in the next session the G coefficient dropped to 0.401. In order to obtain a G coefficient of approximately 0.80, there would have had to be 8-9 questions in the first series and more than 15 questions in the last.

DISCUSSION

We found that a combination of a structured application form and a structured interview had a high acceptability among applicants and assessors alike.

We used a behaviour-based approach in these questions and ratings based on the applicants' ability to reflect on their skills and on their need to improve. By testing the reflective capacity, it is assumed that you can qualify the information used in the selection process [14]. We found that the structured application form and the interview were useful tools in the rating. But they were not sufficient for the assessment of all the key

roles, and documented observed performance from the 6-month introduction training was seen as a useful supplement. A review of psychological studies on structured interviews [10] described that an improved interview structure may also result in increasingly negative applicant reactions. We assume that the individualised part of the application form may have eased the uneasiness concerning a too fixed structure.

Structured interviews are perceived more positively by the users if sufficient information is given to participants in advance [15]. All our applicants were informed about the interview in a letter describing the procedure and the rating, and they found that this information was essential for the observed high acceptability.

It has been shown that candidates in the UK perceived their assessment method to be fair [9]. The participants in our study also perceived the selection process to be very fair, and this may be owing to the personalised guide, which allowed the applicant to go into details about their personal reflections.

It has been shown that structured interviews can produce an acceptable reliability in the recruitment of doctors for psychiatrist training [16].

We demonstrated an acceptable Cronbach's alpha for all interviewers individually and in total, whereas the G coefficient was only acceptable in the first round and not in the second. The reason for this remains unclear. Adding more questions to the interview guide might increase the G coefficient.

There is no gold standard for finding the best applicant. Other endpoints used for satisfactory selection are low drop-out rates and a low number of poor performers during the subsequent specialist training. These problems are limited in Denmark.

Our focus on the applicants' and panellists' impression of an acceptable, fair and reliable selection procedure has demonstrated that our method is a useful, but not the only tool that may be used to find the best trainable applicants.

Strengths and weaknesses

This study includes reflections from both applicants and panellists and, furthermore, one of the panellists was a trainee representative. It uses an "action research-like approach" in the development of a new selection tool. It is a small study though. We cannot ensure data saturation and it is difficult to apply a theoretical perspective to these types of data. Furthermore, we have no data to determine whether the selected applicants perform better than earlier applicants.

It was disappointing that we could not ensure an acceptable G coefficient, and the reliability of the selection procedure could, therefore, be questioned. Consequently, the results should be perceived as important experiences rather than as data with a documented high validity.

Admission interview with applicant for specialist training.

CONCLUSION

A semi-structured personal interview combining individualised elements in the application with standardised behaviour-based questions provided a high degree of acceptability. We were unable to demonstrate a high generalisability, but found an acceptable reliability. The method was found to be feasible and useful in the selection of doctors for specialist training in family medicine provided it is combined with work-based assessment. This view was shared by both panellists and applicants. Our method is now fully implemented in the Region of Southern Denmark. Further studies that include dropout rates are needed to compare our model to other Danish selection models, e.g. the MMI.

CORRESPONDENCE: *Jesper Hesselbjerg Isaksen*, Søkildevej 31, 5700 Svendborg, Denmark. E-mail: jesper.isaksen@dadlnet.dk

ACCEPTED: 20 June 2013

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk.

ACKNOWLEDGEMENTS: We wish to acknowledge the assistance of *Anne Vibeke Schiødt* and *Daniel Roosen* for conducting the interviews and performing data triangulation. We also wish to thank *Jonna* and *Brian Creedon* for having revised our manuscript.

LITERATURE

1. Patterson F, Ferguson E, Lane P et al. A competency model for general practice: implications for selection, training, and development Br J Gen Pract 2000;50:188-93.
2. Patterson F, Ferguson E, Norfolk et al A new selection system to recruit general practice registrars: preliminary findings from a validation study BMJ 2005;330:711-4.
3. Schmidt FL, Hunter JE. The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psych Bull 1998;24:262-74.
4. Robertson IT, Smith M. Personnel selection. J Occup Organ Psych 2001;74:441-72.
5. Randall R, Davies H, Patterson F et al. Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. Arch Dis Child 2006:91;444-8.
6. Dore KL, Kreuger S, Ladhani M et al. The reliability and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate admissions. Acad Med 2010;85(Suppl 10):60-3.

7. www.dsam.dk/flx/english/quality_development/ (Jun 30 2006).
8. Foster C, Godkin L. Employment selection in health care: the case for structured interviewing. HCMR 1998;23:46-51.
9. Wiesner WH, Cronshaw SF. A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. J Occup Psych 1988;61:275-90.
10. Posthuma RA, Morgeson FP, Campion MA. Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. Pers Psychol 2002;55:1-81.
11. Kennedy TJ, Lingard LA. Making sense of grounded theory in medical education. Med Educ 2006;40:101-8.
12. Holstein B. [Triangulation, method and validity]. In: Lunde IM, Ramhøj P. eds. [Art science within health science]. Copenhagen: Akademisk Forlag, 1995:329-38 (Danish).
13. Brennan RL. Generalizability Theory. New York, Springer, 2001.
14. Strasser PB, Improving applicant interviewing – using a behavioral-based questioning approach. AAOHN 2005;53:149-51.
15. Kohn LS, Dipboye RL. The effects of interview structure on recruiting outcomes. J Appl Soc Psych 1998;28:821-43.
16. Rao R. The Structured Clinically Relevant Interview for Psychiatrists in Training (SCRIPT): a new standardized assessment tool for recruitment in the UK. Acad Psychiatry 2007;31:443-6.