Implementing standard setting into the Conjoint MAFP/FRACGP Part 1 examination – Process and issues

Chan SC, Mohd Amin S, Lee TW

Chan SC, Mohd Amin S, Lee TW. Implementing standard setting into the Conjoint MAFP/FRACGP Part 1 examination – Process and issues. *Malays Fam Physician*. 2016;11(2 & 3);2–8.

Keywords:

Standard setting, modified Angoff method, Conjoint MAFP/FRACGP examination

Authors:

Chan Sook-Ching

(Corresponding author)
FAFPM, FRACGP
Community Based Department,
Faculty of Medicine, Universiti Kuala
Lumpur Royal College of Medicine
Perak, Ipoh, Perak, Malaysia.
E-mail: drscchan@gmail.com/
scchan@unikl.edu.my

Sheikh Mohd Amin

FAFPM, FRACGP
Academy of Family Physicians of
Malaysia, Kuala Lumpur, Malaysia.
E-mail: sheikhamin72@amail.com

Lee Tong-Weng

FAFPM, FRACGP Klinik TW Lee, Ipoh, Perak, Malaysia. E-mail: drleetongweng@gmail.com

Abstract

Background: The College of General Practitioners of Malaysia and the Royal Australian College of General Practitioners held the first Conjoint Member of the College of General Practitioners (MCGP)/Fellow of Royal Australian College of General Practitioners (FRACGP) examination in 1982, later renamed the Conjoint MAFP/FRACGP examinations. The examination assesses competency for safe independent general practice and as family medicine specialists in Malaysia. Therefore, a defensible standard set pass mark is imperative to separate the competent from the incompetent.

Objective: This paper discusses the process and issues encountered in implementing standard setting to the Conjoint Part 1 examination.

Discussion: Critical to success in standard setting were judges' understanding of the process of the modified Angoff method, defining the borderline candidate's characteristics and the composition of judges. These were overcome by repeated hands-on training, provision of detailed guidelines and careful selection of judges. In December 2013, 16 judges successfully standard set the Part 1 Conjoint examinations, with high inter-rater reliability: Cronbach's alpha coefficient 0.926 (Applied Knowledge Test), 0.921 (Key Feature Problems).

Introduction

The College of General Practitioners of Malaysia was established in 1973. In 1979, it held its first membership examinations leading to the award of the Member of the College of General Practitioners (MCGP). This was the first higher professional examination in the discipline of general practice/family medicine in Malaysia. Since 1982, with the participation of the Royal Australian College of General Practitioners (RACGP), the examination became the Conjoint MCGP/Fellow of Royal Australian College of General Practitioners (FRACGP) examination. With the formal recognition of the College in 1996, it was renamed as the Academy of Family Physicians of Malaysia (AFPM), and the examination was renamed as the Conjoint MAFP/FRACGP examinations.1

The Conjoint MAFP/FRACGP examination is a high-stakes postgraduate family medicine examination as successful candidates awarded MAFP/FRACGP are recognised as specialists in family medicine registerable under the National

Specialists Register of Malaysia since 2006. The Part 1 examination consists of Applied Knowledge Test (AKT) with 150 multiple choice questions (70 single best questions and 80 extended matching questions) and Key Feature Problems (KFP) with 26 clinical problems. Questions are contributed by both AFPM and RACGP.

Before the Malaysian Ministry of Health's (MOH's) recognition of family medicine specialists and the setting up of a Malaysian National Specialist Register, the numbers taking the examination were small; mainly general practitioners (GPs) who were interested to improve themselves. However, with the recognition given, there are increasing numbers of candidates, including primary care doctors from MOH.

A standard or cut point in an assessment is a score that divides those who perform well enough to pass and those who do not.² Relative standards that involve ranking are commonly used to select the highest scorers for medical school entry when places are limited. Absolute standards are used for tests of competence, to establish whether

the candidates know enough of the particular subject assessed. Absolute standards using a fixed pass mark is easy to use and can be applied to different forms of assessments. The judges are also comfortable in using it. However, it does not take into consideration the difficulty of examination items. If the test items are very difficult, it is possible for the competent candidates to fail. On the other hand, a very easy assessment will allow less competent candidates to pass. As such, the modified Angoff method has been used to standard set high-stakes examinations.

Prior to standard setting, a fixed passing mark of 66% was used. In keeping with RACGP³ and with the increasing number of candidates, the Board of Examiners (BOE) of AFPM decided in 2011 that standard setting for the Conjoint Examination is imperative for a defensible pass mark. In preparation for the introduction, training sessions were initiated from 2011. The modified Angoff method was selected as it has been successfully used in high-stakes examinations².⁴¬ and is the method used by the RACGP in their examinations.8

In this method, a group of judges must first define the characteristics of a hypothetical borderline candidate, then determine how the borderline candidate would answer each test item and estimates the proportion who would get it correct. Each judge independently standard sets the test items. The group then comes together to discuss each test item. The judges can change their scores. The judges' final estimates are averaged for each item and the sum of these averages is the pass mark or cut score. This paper discusses the process and issues encountered in training and implementation of standard setting into the Conjoint Part 1 examinations using the modified Angoff method.

Methods

The first standard setting training session in 2011 was conducted by an invited external speaker and the first author, who is a senior examiner of the BOE. Both have vast experience in standard setting in the undergraduate programme examinations in the Royal College of Medicine Perak. For most of the participants, this was their first exposure to standard setting and to the modified Angoff method.

For the subsequent workshops, the participants were selected based on age, gender, place of work and willingness to participate in standard setting.^{2,5} Downing et al.⁵ emphasised the important characteristics such as knowledge of target population, content expertise, fairness, open mindedness, unbiased attitude, willingness to follow instructions and understanding the task as judges, and also the importance of balancing the panel of judges with demographic variables such as ethnicity, gender, geography and subspecialisations with practical considerations playing a major role in judge selection. Norcini² suggested that the judges should be a mixture of "teacher, practitioner, generalist and specialist" depending on test purpose and stated the importance of a balance of personal attributes like gender, race and age to avoid real or perceived conflicts of interest.

As such the participants for subsequent workshops were selected from members of the AFPM BOE based on the attributes mentioned above. The majority (76%) were full-time practising family physicians from both the private and public sectors and a minority (24%) were academic family medicine specialists. Their places of practice were from all over Malaysia. They were of both genders, different ethnic groups and of different age groups (ranging from trainee examiners to senior examiners). They formed the pool of judges for the panel of standard setters for the Part 1 Conjoint examination (see **Table 1**).

In 2012, the second workshop concentrated on standard setting AKT. In 2013, the third workshop covered both AKT and KFP.

Prior to the second and third workshops, the participants were given introductory articles on standard setting for prereading to ensure they had some knowledge on standard setting. During the workshop, the participants were briefed on the modified Angoff method and on the characteristics of a borderline candidate, followed by handson exercises with past year's questions. Participants independently standard the past year's AKT and KFP, followed by a group discussion led by a moderator. The judges with the highest and lowest scores for each test item were requested to give the reasons for their respective scores. All the judges were allowed to join the discussion and to alter their standard set scores. The mean of the revised scores of each test item

was summed to get the cut score or passing mark. Reality checks in the form of actual candidates' performance in past year's test items were provided whenever the judges' scores differed widely from the actual performance. This was done not to influence the judges' scores towards the candidates' performance but used for training in situations when the facilitator felt that the judges held unrealistic views/wrong concepts, for example, equating a borderline candidate to an average or distinction candidate or on the other extreme, thinking very lowly of all borderline candidates' performance. Another issue was a tendency for some participants to think of how borderline candidates should perform without appreciating the difference between "will" perform and "should" perform.5 Reality checks are also used by RACGP but as a final step in standard setting their actual examination. In standard setting the actual Conjoint examination, reality checks are not used at all as the standard setting is done simultaneously as the actual examination and the papers have not been marked yet and the results are not known. During the training sessions, the standard setting was done on the past year marked actual examination questions with item analysis done and the percentages of candidates correctly answering all questions known.

At the end of the workshop, the judges were asked to feedback by bringing up any issues encountered. After the third workshop, all felt

that a detailed guideline on the characteristics of the borderline/minimally competent candidate was required. It was decided that after the workshop the judges would continue the discussion via email. Based on their feedback obtained through email communication, a detailed guideline on the characteristics of the borderline/minimally competent candidate was produced and circulated together with a section on how to score AKT. The facilitators of the second and third workshops also sent a written feedback report to the Head (chief examiner) of the BOE.

In December 2013, 16 judges selected from the workshop participants' standard set the Conjoint Part 1 examination, over a period of 2 days. This was held concurrently with the actual Part 1 examination to prevent the leakage of questions. Selection criteria for the judges again encompassed the characteristics such as age, gender, place of work and willingness to participate as mentioned above^{2,5} (see **Table 1**).

The standard set scores of the judges, before and after discussion, were analysed using the Statistical Package for Social Sciences version 17.0 (SPSS Inc, Chicago, IL) for inter-rater reliability (Cronbach alpha coefficient). Variability between the judges' scores was measured by the range in percentage between the highest and lowest total scores of the judges obtained in AKT (for 150 items) and KFP (26 cases).

Table 1. Demographic characteristics of judges in part 1 conjoint exam. December 2013

Characteristics of judges	Number of judges N = 16	Percentage
Place of practice		
i) Full-time practising	12	75
a) in private family practice clinics	9	56
b) in MOH as family medicine specialist	3	19
ii) Academic family medicine (university based)	4	25
Seniority of examiners		
i) Senior examiner	5	31
ii) Examiner	7	44
iii) Trainee examiner	4	25
Gender		
i) Male	9	56
ii) Female	7	44
Ethnic group		
i) Malay	5	31
ii) Chinese	9	56
iii) Indian	2	13
Age group		
i) 50 years and above	10	62
ii) 30–49 years	6	38

Results

Initially, the main issues were lack of experience in standard setting and difficulty grasping the concept and process of the modified Angoff method.

In 2013, the majority (71%) were comfortable with the process but a few still had difficulty conceptualising a borderline/minimally competent candidate. Through email communications among the participants after the 2013 workshop, a detailed guideline including a section on how to score AKT was produced and circulated among the participants. The guideline provided the background information on the Conjoint examinations and standard setting using the modified Angoff method, followed by a definition of the borderline candidate. For the Conjoint examination, the borderline candidate was defined as a "just passed" or "minimally competent" candidate and not as one "sitting on the fence with 50% chance of passing or failing". The former is a higher standard required, as the Conjoint examination is an exit postgraduate examination, whose graduates are deemed as safe enough to practice in independent family practice and minimally competent to be a family medicine specialist in Malaysia. The main section of the guideline described the characteristics of the minimally competent candidate in history taking, communication, physical examination, diagnosis, investigations, and management skills. In the Appendix, guidelines on scoring AKT were given, which included assessing the importance of topic/ commonly encountered conditions, difficulty of questions and number of distractors. Based on these considerations, the judges would enter the proportion (ratio) of minimally competent candidates whom they thought would get the correct answer (see Table 2). For KFP, the judges guided by the characteristics of a minimally competent candidate in the guideline, assessed the number of points (marks) the minimally competent candidate would score in each question. Table 2 shows a sample of questions and the judges' standard set scores before and after discussion. The points given in each question of a case (clinical problem) were multiplied by the allotted weightage and summed to obtain the total score for each case. The maximum score equalled 12 for each case.

In December 2013, 16 judges successfully standard set the Part 1 Conjoint examinations. **Tables 2 and 3** show a sample of the judges' standard set scores before and after discussion. Discussion among the judges after scoring

independently narrowed the variability (range between highest and lowest scores) between judges' total scores from 14% to 11% (AKT) and 24% to 10% (KFP) overall and resulted in high inter-rater reliability with Cronbach's alpha coefficient, which increased from 0.683 (AKT) and 0.810 (KFP) before discussion to 0.926 (AKT) and 0.921 (KFP) after discussion.

Discussion

Identified as critical to success in standard setting were 1) the judges' understanding of the process of the modified Angoff method, 2) defining the characteristics of the borderline candidate and 3) selection of judges.

1) The judges' understanding of the process of the modified Angoff method

The above issue was overcome by repeated workshops with hands-on exercises. The "judges" performance in the hands-on exercises was informally evaluated by senior board members experienced in standard setting. A general feedback was given as part of the discussion at the end of the workshop.

2) Defining the characteristics of the borderline candidate.

Participants initially had their own idea of the characteristics of a borderline candidate, some visualising him as an average GP in Malaysia, or as a senior medical officer in MOH. A few felt the borderline candidate should be excellent in most aspects of management.

Introducing the RACGP's guidelines on the characteristics of borderline or minimally competent candidate was very helpful.8 An important issue in Malaysia was that those who passed must not only be competent safe family physicians who can work independently but must be minimally competent to function as a family medicine specialist in MOH. Previously, the Conjoint examination was aimed at certifying GPs in the private sector as safe family physicians. Since the introduction of the National Specialist Register in Malaysia and the recognition by MOH, those who passed are accepted family medicine specialist after credentialing. Family medicine specialists in MOH have additional responsibilities of

Table 2. Judges' standard setting scores (before and after discussion) in a sample of questions from AKT paper

Item		Judge number and standard set scores before discussion															Range	Item
No	J1	J2	J3	J4	J5	J6	J 7	J8	J9	J10	J11	J12	J13	J14	J15	J16	(%)	mean
1	0.70	0.60	0.60	0.80	0.60	0.60	0.60	0.70	0.70	0.70	0.70	0.80	0.80	0.65	0.80	0.65	20	0.69
2	0.60	0.60	0.50	0.70	0.70	0.60	0.70	0.70	0.80	0.70	0.80	0.65	0.80	0.55	0.70	0.65	30	0.67
3	0.60	0.70	0.70	0.80	0.70	0.80	0.55	0.50	0.60	0.60	0.55	0.50	0.60	0.50	0.50	0.65	30	0.62
4	0.70	0.80	0.60	0.80	0.80	0.80	0.55	0.70	0.70	0.70	0.70	0.65	0.50	0.70	0.70	0.70	30	0.69
5	0.60	0.70	0.50	0.80	0.60	0.60	0.60	0.60	0.70	0.70	0.65	0.50	0.55	0.65	0.80	0.70	30	0.64
6	0.70	0.80	0.80	0.90	0.80	0.70	0.60	0.70	0.80	0.70	0.65	0.80	0.60	0.75	0.60	0.75	30	0.73
7	0.80	0.70	0.70	0.90	0.80	0.65	0.70	0.70	0.70	0.70	0.80	0.70	0.70	0.65	0.90	0.75	25	0.74
8	0.50	0.70	0.70	0.80	0.80	0.50	0.70	0.50	0.40	0.50	0.55	0.40	0.60	0.65	0.40	0.70	40	0.59
9	0.50	0.70	0.70	0.70	0.60	0.65	0.65	0.60	0.60	0.70	0.60	0.80	0.70	0.80	0.80	0.75	30	0.68
10	0.50	0.70	0.60	0.60	0.50	0.40	0.55	0.40	0.50	0.60	0.50	0.50	0.45	0.40	0.50	0.65	30	0.52
Total	6.20	7.00	6.40	7.80	6.90	6.30	6.20	6.10	6.50	6.60	6.50	6.30	6.30	6.30	6.70	6.95	17	0.657*

Item	Judge number and standard set scores after discussion												Range	Item				
No	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12	J13	J14	J15	J16	(%)	mean
1	0.70	0.60	0.60	0.80	0.60	0.60	0.60	0.70	0.70	0.70	0.70	0.80	0.80	0.65	0.80	0.65	20	0.69
2	0.60	0.60	0.60	0.70	0.70	0.60	0.70	0.70	0.75	0.70	0.75	0.65	0.70	0.55	0.75	0.65	20	0.67
3	0.60	0.70	0.70	0.75	0.70	0.70	0.55	0.50	0.60	0.60	0.55	0.50	0.60	0.50	0.50	0.65	25	0.61
4	0.70	0.80	0.60	0.80	0.80	0.80	0.60	0.70	0.70	0.70	0.70	0.65	0.60	0.70	0.70	0.70	20	0.70
5	0.60	0.70	0.60	0.75	0.60	0.60	0.60	0.60	0.70	0.70	0.65	0.60	0.60	0.65	0.75	0.70	15	0.65
6	0.70	0.80	0.80	0.80	0.80	0.70	0.60	0.70	0.80	0.70	0.65	0.80	0.60	0.75	0.60	0.75	20	0,72
7	0.80	0.70	0.70	0.80	0.80	0.65	0.70	0.70	0.70	0.70	0.80	0.70	0.70	0.65	0.80	0.75	15	0.73
8	0.50	0.70	0.70	0.75	0.70	0.50	0.70	0.50	0.50	0.50	0.55	0.65	0.60	0.65	0.50	0.70	25	0.61
9	0.60	0.70	0.70	0.70	0.60	0.65	0.65	0.60	0.60	0.70	0.60	0.80	0.70	0.80	0.80	0.75	20	0.68
10	0.50	0.60	0.60	0.60	0.50	0.40	0.55	0.40	0.50	0.60	0.50	0.50	0.45	0.40	0.50	0.60	20	0.51
Total	6.30	6.90	6.60	7.45	6.80	6.20	6.25	6.10	6.55	6.60	6.45	6.65	6.35	6.30	6.70	6.90	13	0.657*

^{*} Obtained by averaging sum of all item means.

Table 3. Judges' standard setting scores (before and after discussion) in a sample of questions from KFP paper

Item						Judge nu	mber and	l standard	l set score	s before d	liscussion						Range (%)	Case mean
No	J1	J2	J3	J4	J5	J6	J 7	J8	J9	J10	J11	J12	J13	J14	J15	J16		
Case 1																		
Q1	3	3	3	3	3	3	3	4	3	4	4	3	2	3	3	3	-	-
Q2	3	3	2	2	3	3	3	3	2	4	3	3	3	3	3	3	-	-
Q3	4	6	5	4	5	6	6	6	5	6	7	4	4	6	5	6	-	-
Total 12	7.72	8.65	7.28	6.82	8.18	8.65	8.65	9.7	7.28	10.60	10.17	7.72	6.67	8.65	8.18	8.65	33	8.35
Case 2	Case 2																	
Q1	3	3	3	3	3	3	3	3	3	3	4	3	3	2	3	3	-	-
Q2	3	3	3	3	2	2	2	4	4	2	4	3	3	2	3	3	-	-
Total 12	7.92	7.92	7.92	7.92	6.48	6.48	6.48	9.36	9.36	6.48	10.56	7.92	7.92	5.28	7.92	7.92	44	7.74
Case 3																		
Q1	2	2	3	1	2	3	3	2	3	3	3	3	3	2	2	0	-	-
Q2	2	2	2	4	2	4	4	3	4	2	2	2	2	2	3	0	-	-
Q3	3	3	3	3	3	3	3	3	1	3	4	3	2	3	3	2	-	-
Q4	5	6	4	6	5	7	6	6	6	5	7	5	5	5	6	5	-	-
Total 12	6.57	7.10	6.03	7.40	6.57	9.13	8.60	7.55	7.40	7.17	8.83	6.57	7.17	6.57	7.55	3.87	44	7.13
Total of 3 cases	22.21	23.67	21.23	22.14	21.23	24.26	23.73	26.61	24.04	24.25	29.56	22.21	21.76	20.50	23.65	20.44	25	7.739*

Judge number and standard set scores after discussion Range Item Case No J2 J5 J8 J9 J10 J16 mean Case 1 Q1 3 2 2 Q2 3 3 3 3 3 3 3 4 3 3 3 3 3 3 Q3 4 6 5 5 5 6 6 5 5 5 3 4 5 6 5 6 Total 7.28 7.72 8.65 7.28 8.18 8.18 8.65 8.65 8.18 9.08 7.25 7.13 8.65 8.18 8.65 16 8.09 12 Case 2 Q1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 3 3 3 3 3 3 Q2 3 3 2 3 3 7.92 7.92 7.92 7.92 6.48 6.48 6.48 7.92 7.92 6.48 9.12 7.92 7.92 7.92 7.92 7.63 12 Case 3 Q1 2 2 2 3 2 Q2 2 2 2 4 2 4 4 3 4 2 2 2 2 2 3 2 4 Q3 3 3 3 3 3 3 3 3 1 3 3 3 3 2 Q4 5 6 4 6 5 7 6 6 6 7 5 5 6 Total 7.10 6.63 6.57 7.55 7.40 8.83 7.17 6.57 7.03 7.40 22.21 23.67 21.83 23.5 21.23 24.26 23.73 23.65 22.6 22.73 25.20 22.81 22.22 23.14 23.65 23.6 11 7.709*

Table 3. Judges' standard setting scores (before and after discussion) in a sample of questions from KFP paper (Cont'd)

supervising and seeing referred cases from the medical officers in their clinics and running specialised clinics for patients with special problems such as HIV/AIDS. So, the workshop participants were asked to focus on both of these aspects when they standard set AKT and KFP. The detailed guideline on the characteristics of a borderline/minimally competent candidate together with the section to aid in marking AKT was found to be very helpful by the participants.

3) Selection of judges

The selection of the panel of judges in standard setting is very important as it has a definite impact on the cut-off score. As mentioned earlier, the judges selected included a mixture of senior to junior examiners from the AFPM BOE. All the BOE members had themselves passed the Conjoint examination and had gone through the AFPM's process of examiner selection. They were all content experts and knowledgeable of the curriculum and the examination process. All had experience in teaching undergraduate or postgraduate students either as academics

in medical faculties or mentors in the AFPM. They had demonstrated integrity and ability to work as a team. The judges were of both gender, different ethnic groups and practice background.^{2,5} They were open minded and willing to devote their time to the task. The authors felt that diversity was needed to increase the credibility of the established cut-off score. It is conceivable that age, gender and place of work may have an impact on the judges' decision on an item and greater diversity would appear to be fairer to average out differences.

In conclusion, inexperience and difficulty in grasping the concepts of standard setting and borderline candidate were successfully overcome by repeated hands-on training, provision of detailed guidelines and careful selection of judges. Standard setting using the modified Angoff method showed high inter-rater reliability with Cronbach's alpha coefficient of 0.926 for AKT and 0.921 for KFP after discussion. Standard setting has been successfully implemented in the Conjoint Part 1 examination (a postgraduate family medicine examination) from December 2013.

^{*} Obtained by averaging sum of all case means.

Acknowledgement

We would like to thank the Chief Examiner of the Board of Examiners for the permission to have this manuscript published and the members of the Academy of Family Physicians of Malaysia who participated in the training and standard setting sessions of the Part 1 Conjoint examination. Also, special thanks to AP Dr. R Rajasingam Shanmugam for his advice and suggestions on standard setting.

How does this paper make a difference in general practice?

- This paper aims to create awareness of the Conjoint MAFP/FRACGP examinations to our colleagues in general practice.
- The Conjoint examinations not only assess competence for safe independent general practice but also provide a pathway for recognition as a specialist in family medicine in Malaysia.
- Previously fixed absolute standards were used to determine the passing mark but since 2013, standard setting has been introduced to ensure a defensible passing mark separates the competent from the incompetent, in line with the RACGP requirements.

References

- Conjoint MAFP/FRACGP Examination.
 Handbook for Candidates, Mentors &
 Examiners. Year 2009. Academy of Family
 Physicians of Malaysia.
- 2. Norcini JJ. Setting standards on educational tests. *Med Educ.* 2003;37(5):464–9.
- Jones R, Rawlin M, Atkinson K, et al. Lessons from the past. Historical trends in the RACGP examinations. *Aust Fam Physician*. 2009;38(9):708–10.
- Fowell SL, Fewtrell R, MClaughlin PJ.
 Estimating the minimum number of judges required for test centred standard setting on written assessments. Do discussion and iteration have an influence? Adv Health Sci Educ Theory Prac. 2008;13(1):11–24.
- Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med.* 2006;18(1):50–7.
- Barman A. Standard setting in student assessments: is a defensible method yet to come? *Ann Acad Med Singapore*. 2008;37:957–63.

- Davis MH, Ponnamperuma G. Standard setting. Theme: Assessment (for postgraduate certificate, postgraduate diploma, masters in medical education courses). Centre for medical education. *Dundee*. November 2006.
- Egglestone D. Standard setting for Fellowship examination. 2010 & 2011.
 Slide presentations produced for the Royal Australian College of General Practitioners.