

5. Monitoring IELTS Examiner Training Effectiveness

A preliminary study

*Clare McDowell
Australian Research & Testing Services*

Publishing details

**International English
Language Testing System (IELTS)
Research Reports 2000
Volume 3**

Editor: Robyn Tulloh

IELTS Australia Pty Limited
ABN 84 008 664 766
Incorporated in the Australian Capital Territory
Web: www.ielts.org

© 2000 IELTS Australia.

This publication is copyright. Apart from any fair dealing for the purposes of private study, research or criticism or review, as permitted under the Copyright Act, no part may be reproduced by any process without written permission. Enquiries should be made to the publisher.

National Library of Australia
Cataloguing-in-Publication Data
2000 ed
IELTS Research Reports 2000 Volume 3
ISBN 0 86403 036

5. Monitoring IELTS Examiner Training Effectiveness

A preliminary study

*Clare McDowell
Australian Research & Testing Services*

1.0 Background to the Research Project

1.1 History of IELTS Examiner Training

The key to reliability and hence validity of a direct or semi direct test is effective and internationally consistent examiner training. Since the launch of the IELTS test in 1989, the vast majority of examiners operating in Australia and at IDP centres off-shore have been trained in the face to face mode using materials initially created in 1989 and updated in subsequent years. However, these materials were designed to be used both in a self-access mode as well as training sessions and most of the examiners at The British Council centres and in the United Kingdom did their training through the self-access mode. Monitoring has shown that this dual training model inevitably led to some inconsistencies in rating particularly where the interview was concerned.

1.2 The New Training Package

With the introduction of the revised version of IELTS in 1995, a decision was made to train all new examiners wherever possible through face to face training and to produce a new training package to this end. The training package was developed in Australia and the United Kingdom by two senior examiners and consists of sets of officially rated scripts for use in the Writing training session and video taped interviews accompanied by explanatory profiles for the Speaking session together with extensive notes to assist the trainers based on grades awarded by a number of senior examiners worldwide.

The new training package was launched in May 1997 in Sydney when Senior Examiners from IELTS Australia centres were given an introduction to the materials and it has now been used to train new IELTS examiners both at The British Council centres as well as IELTS Australia centres in a number of countries including the United Kingdom, Australia, New Zealand, the Indian sub-continent and South East Asia.

1.3 Benefits of Face to Face Training

Face to face training has the immediate effect of delivering to the participants a standard approach to the procedures involved in rating scripts and, perhaps more importantly, to conducting and rating interviews. This contrasts with the self-access mode which is open to variations of interpretation. It is therefore the preferred mode of training but is dependent on there being an effective and user-friendly package available.

2.0 Aims of the Research

The main thrust of this research was to survey trainers and participants recently involved in an examiner training session where the new training materials were used to find out how they felt about the training itself and whether or not the materials were perceived both by the participants and the trainers to be effective.

A subsidiary aim was to end up with an effective post training questionnaire that could be delivered on a regular basis by trainers to gather the opinions of the participants in order to improve and standardise training worldwide.

A third aim was to find out how consistently the training materials are being used and if and when trainers felt the need to adopt a different approach, to ascertain how these innovations could be incorporated into the mainstream training if appropriate through a process of regular reporting back.

3.0 Methodology

The following approach was adopted:

- Design and trial the two questionnaires known as the Trainer Questionnaire and the Participant, or Trainee, Questionnaire.
- Distribute the questionnaires via the trainers to centres where IELTS training had recently taken place using the new training package.
- Collate the data received back.
- Analyse the data and draw conclusions from the responses received.
- Make recommendations with regard to the training package and make the questionnaires or amended versions thereof available for inclusion in the training package.
- Follow up with ongoing reporting after all new training sessions.

4.0 The Surveys

4.1 Aims of the Questionnaires

The questionnaires were designed to be precise and able to be completed in a time frame of under five minutes. Once the design was finalised after a brief trial on three senior examiners, the questionnaires (Appendix 5.1 and 5.2) were printed on two sides of one page, for ease of distribution and handling.

The trainer survey contained 30 items and the participant survey contained 35 items. Respondents were asked to rate on a scale of 1- 5 a number of aspects of the training session and training package. Some items were open questions which allowed respondents to include their views in their own words.

The survey of participants aimed to find out:

- how they felt about the training format and physical conditions
- how they felt about the writing session
- how they felt about the speaking session
- how well prepared they felt before undertaking the certification procedure
- their general impressions and opinions of the training procedure.

The survey of trainers aimed to find out:

- how they felt about the training format and materials generally
- how they felt about the scripts and accompanying notes
- how they felt about the videos and accompanying notes
- how well they felt the materials were able to mould the participants
- their general impressions and opinions of the package.

4.2 Gathering the Data

After preparing and trialing the drafts, the final questionnaires were drawn up and a letter despatched to the trainers who had conducted training recently. A decision was made to distribute the questionnaires via the trainers so that trainers could be involved and would not feel excluded. A letter explaining the procedure was sent to the trainers. Where no response was forthcoming, the direct approach was adopted with questionnaires being sent to the participants. Where training had taken place more than six months earlier, it was decided not to contact these examiners as the time lapse was too great for an accurate response. In addition, the United Arab Emirates which would have yielded many responses, was excluded from the batch as it turned out that the trainer had not used the new training package.

4.3 Centres Contacted

The following centres were contacted through the senior examiners and/or trainers who had conducted training and responses received as set out below:

- | | |
|--------------------------|-----------------------------------|
| • Phnom Penh via Bangkok | Trainer response only |
| • Taipei | Trainer and participant responses |
| • Bali | Trainer and participant responses |
| • Surabaya | Trainer and participant responses |
| • Jakarta | Trainer and participant responses |
| • Sarawak | Trainer and participant responses |

- India Own survey returned – no responses
- Karachi Trainer response only
- Adelaide Trainer and participant responses
- Melbourne Trainer and participant responses
- Auckland Trainer and participant responses – own survey returned also
- Perth Trainer and participant responses

All participants were given the opportunity to remain anonymous though none took up this offer. The overall response has been somewhat disappointing with 45 participants responding, though the responses so far received have been very thorough and helpful. Two trainers have supplied their own data from surveys they conducted themselves but the data could not be included here other than in the form of overall comments.

The surveys should ideally have been given to the participants immediately after the training sessions, but the timing of this research project made that impossible. Nevertheless, those who responded managed to do so effectively and a good cross section of responses has been received from a variety of training situations. As more are received, they can be added to this collection of data as an addendum to this report.

5.0 The Data and Findings

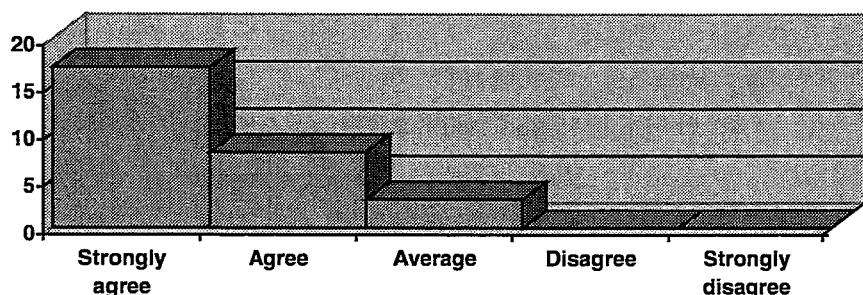
5.1 The Post-Training Participant Questionnaire

Respondents were asked to rate the following propositions from 1 to 5 with 1 being '*strongly agree*' and 5 being '*strongly disagree*'. In some cases, no response was provided.

General Comments

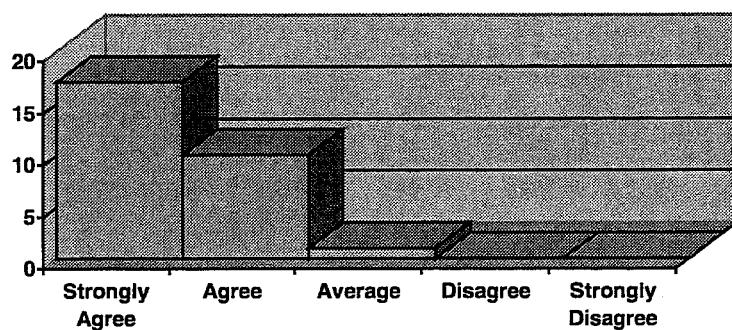
Q 1 The training was well organised.

All but a couple of the respondents felt that their training sessions had been well organised and worthwhile.



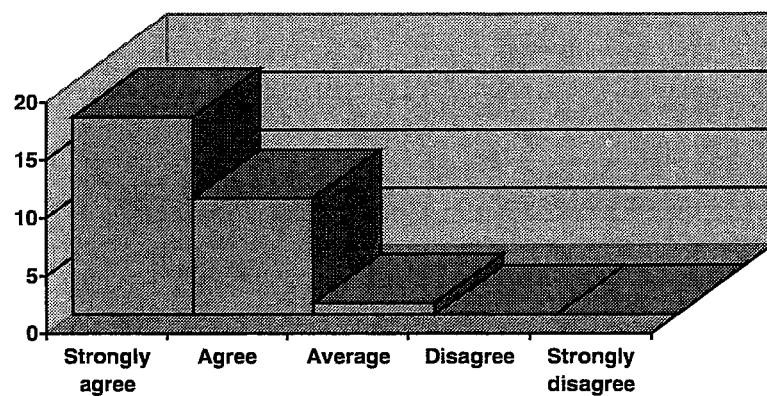
Q 2 The room used was satisfactory.

The training was carried out in suitable venues though some participants felt that conditions for the mock interviews had not been ideal. Space is always a critical issue in training centres but poor conditions can severely upset the training.

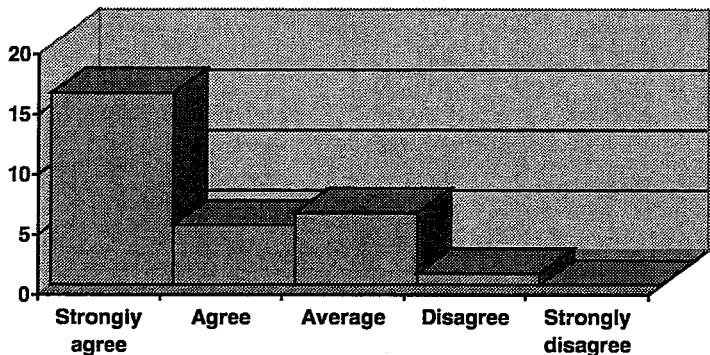


Q 3 The two day format is appropriate.

Some participants reported that they had only been able to spend one day of training which is not recommended procedure at all. One centre had offered training with a day's break in between the sessions, and this had been welcomed. There is a need to ensure that if training is to take place, then the time guidelines are adhered to.

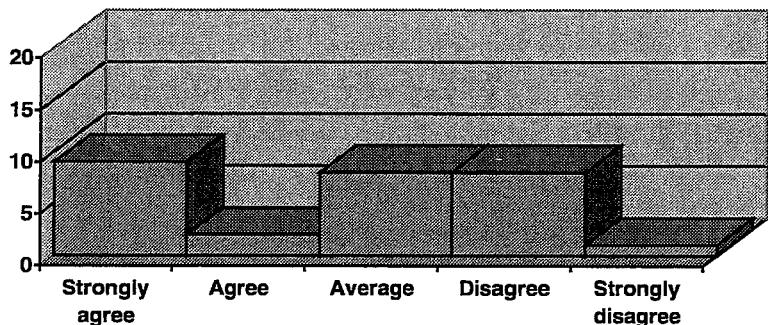


Q 4 All participants were able to participate fully in the activities.



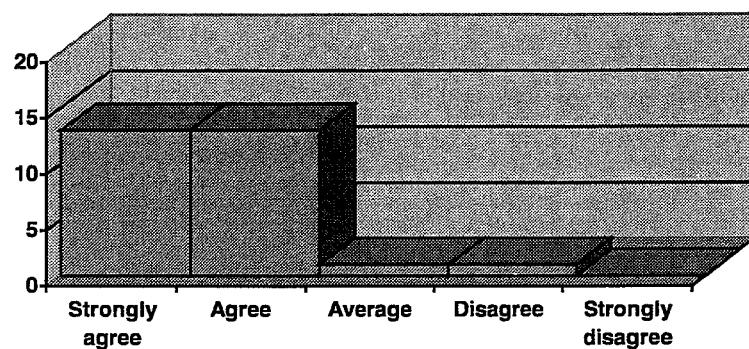
It seems that trainers were able to involve their participants satisfactorily and ensure good participation.

Q 5 The relationship between the three parties (University of Cambridge Local Examinations Syndicate, IDP Education Australia and The British Council) was clearly explained.



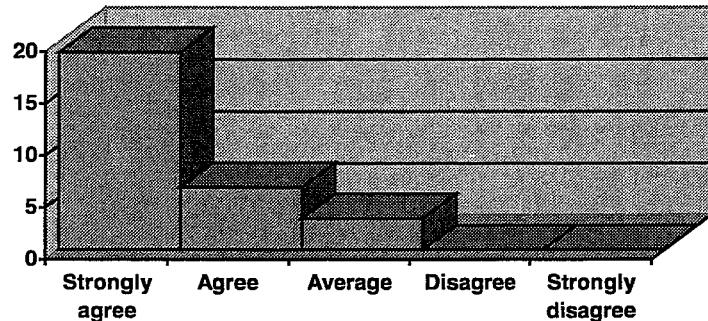
The responses to this question clearly showed participants did not leave their training session with a clear idea of how the three parties are mutually linked. More information on this would help to clear up some regular misunderstandings about the roles of the parties, particularly with regard to the development and delivery of the tests.

Q 6 An overview of the whole IELTS test was provided.



This aspect of the test is well covered in the training materials and the message has obviously been clearly imparted.

Q 7 The examiner's role in relation to the writing and speaking was clearly explained.

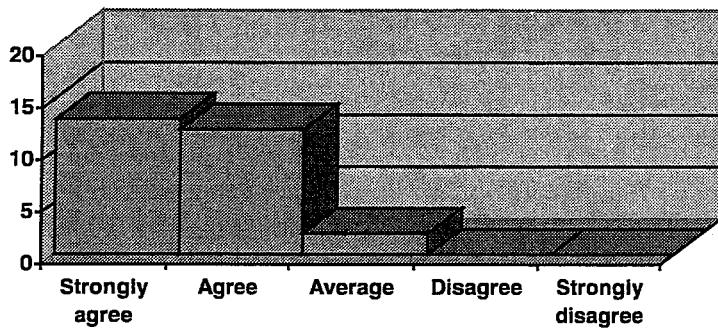


Most participants indicated that their role had been clearly articulated. Prior to training, they may not be aware of the difference between the direct and indirect aspects of the IELTS test and the implications of this on the way the test is marked and thus on test reliability.

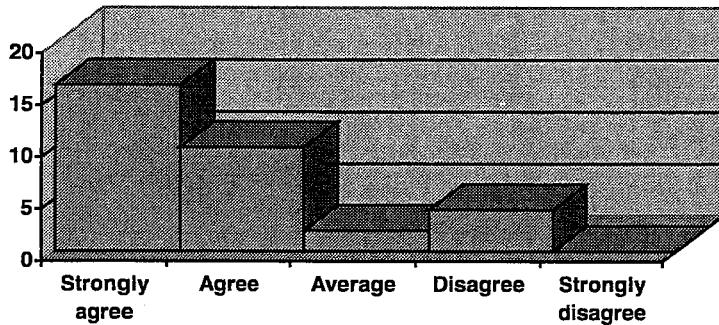
Day One – The Writing Module

Q 8 The aims of the IELTS writing module were clearly outlined.

There was a high degree of agreement here.

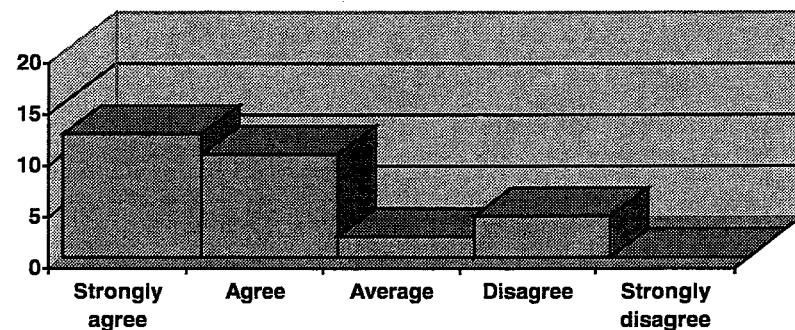


Q 9 There was a clear separation between Writing Task 1 and Task 2.



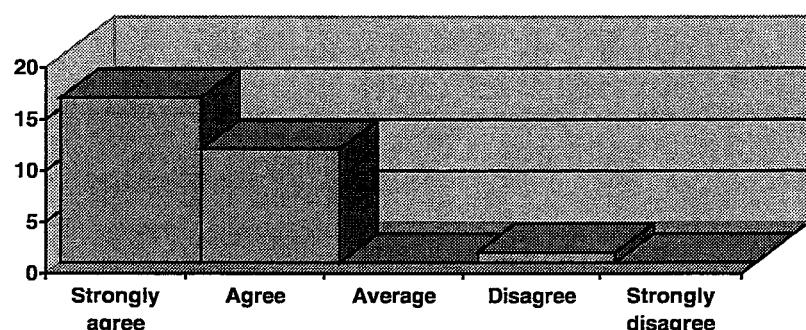
It is important that trainee examiners appreciate the different writing skills being tested in the two writing tasks as this affects the marking procedure. This is best achieved by making a clear separation between the two task types in the training. This may not have been done in all cases.

Q 10 Adequate time was allowed to 'digest' the band descriptors.



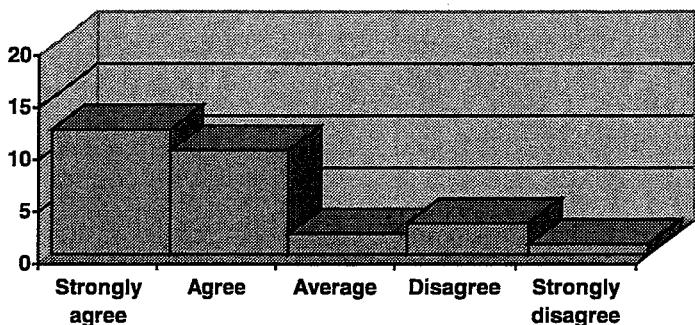
In most cases, there appears to have been ample time given to allow the participants to take in the descriptors, though again some did not agree.

Q 11 The profile and global rating methods were both explained.



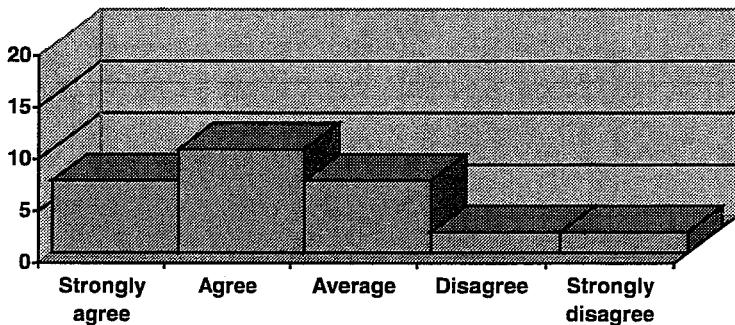
This aspect of the training has obviously been dealt with thoroughly. It is important that trainers offer both methods from the start, without necessarily insisting on one over the other.

Q 12 There were sufficient scripts supplied to exemplify the levels.



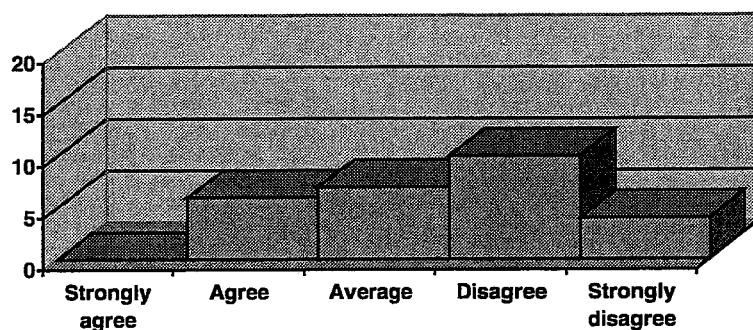
Most respondents felt that there were enough scripts to deal with the levels. Those who disagreed would have liked to see some higher scripts in the Academic writing.

Q 13 There were sufficient scripts provided to exemplify problem areas.



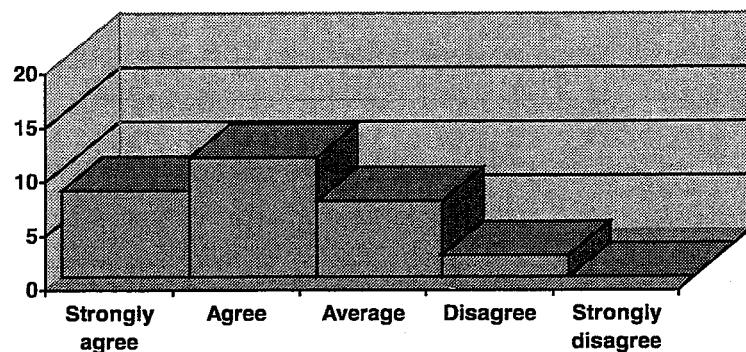
This may be an area of the training materials that needs review. However, the recent addendum to the materials concerning short scripts and scripts off the topic may be sufficient to deal with this aspect of the training.

Q 14 I had difficulty grasping the IELTS levels.



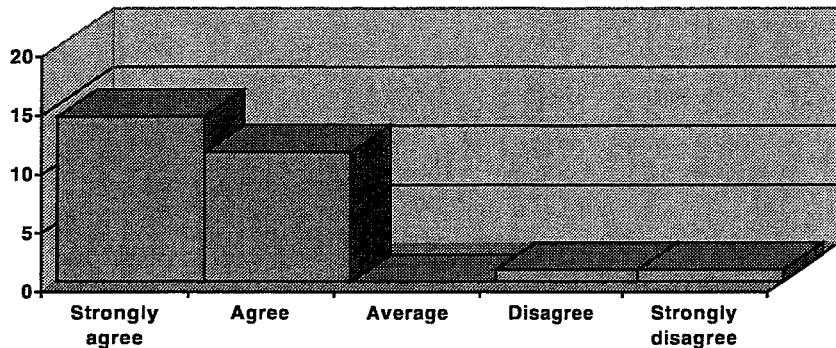
Most respondents did not feel that they had experienced difficulty grasping the levels. This question had been included, however, because in previous training sessions participants had often indicated surprise at the way in which the lower bands operate.

Q 15 By the end of the day I felt I was in tune with the rating procedure.



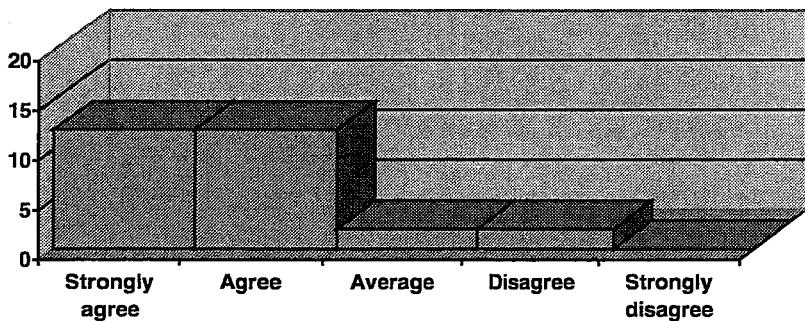
There was an apparent spread of opinions on this question. Some people commented that they had not had enough time to grasp everything or that they needed more practice, but the majority seemed positive.

Q 16 I benefited from the feedback on the supplementary 'homework' scripts.

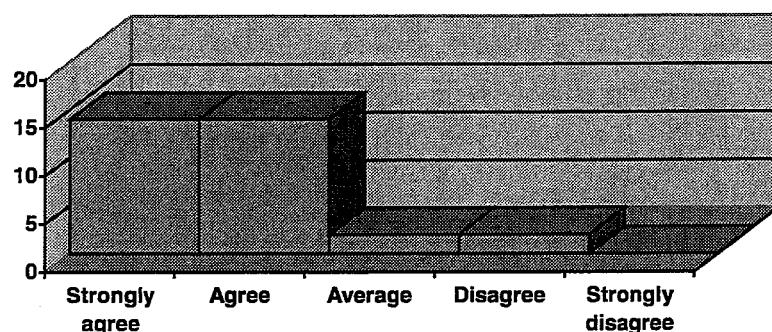


The concensus of opinion was that the additional supplementary scripts constituted a helpful addition to the training package. It was also encouraging to see that everyone had taken advantage of them. There may be an argument for adding to this set and thereby including some more 'problematic' scripts or levels requested.

Q 17 I felt adequately prepared to undertake the certification scripts.



Most people felt ready to do their certification scripts and no negative comments were received. The certification procedure is seen by most participants to be an essential if not always pleasant, part of the procedure.

Q 18 The course prepared me well to work as an IELTS examiner.

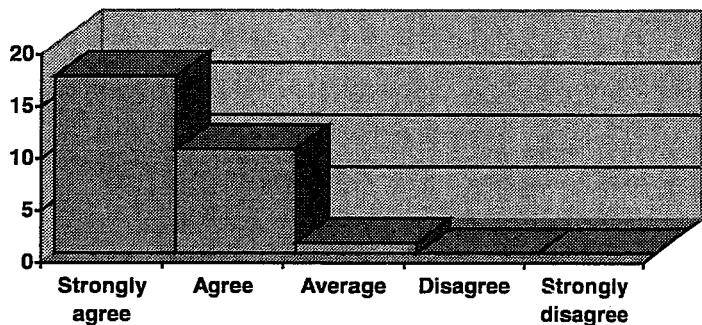
This question was intended to find out to what extent the participants felt equipped to be IELTS examiners after doing the course. Clearly the majority felt that the requirements of an examiner had been spelt out. It was also a reminder that training needs to include information about security issues and the responsibilities of confidentiality and willingness to work as an examiner once one is familiar with the procedure. The difference between training and providing information about the test needs to be made clear.

Q 19 Which aspects of the training were the most useful?

- *Working through the scripts with other participants and then discussing with the trainer*
- *Discussions with colleagues/ comparing ratings*
- *Getting the trainer's view/ interpretation of the descriptors*
- *Homework scripts and feedback.*

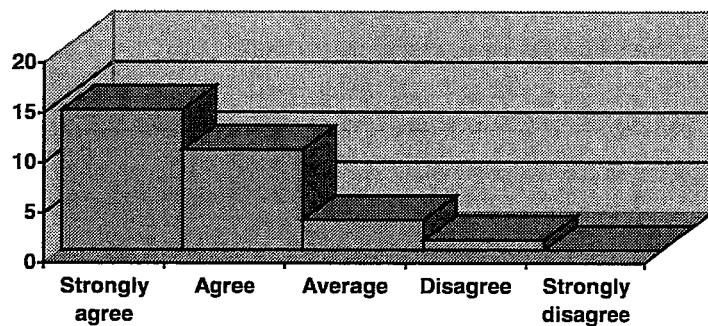
Day Two – The Speaking Module

Q 20 The aims of the IELTS speaking sub-test were clearly outlined.

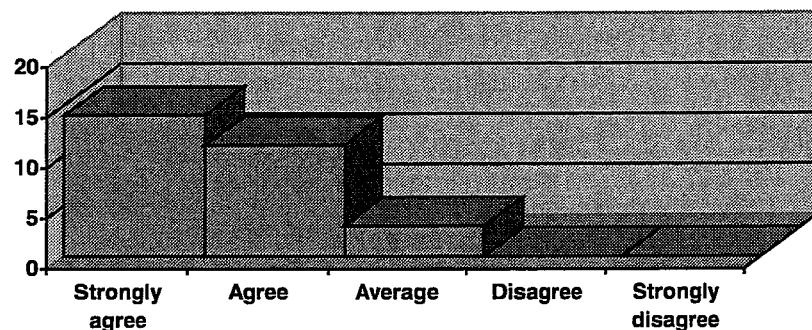


This aspect of training was obviously well covered in all cases.

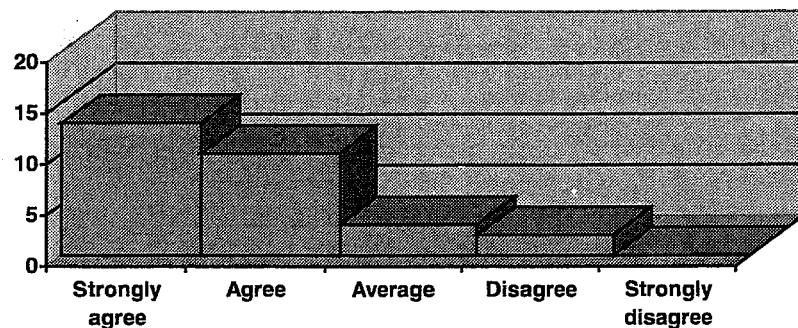
Q 21 The dual role of the examiner was discussed.



Most people indicated that this had been dealt with. Reflecting on this aspect of the work was intended to remind the examiners that being both interlocutor and assessor carries with it a number of responsibilities and draws on more than one examining skill.

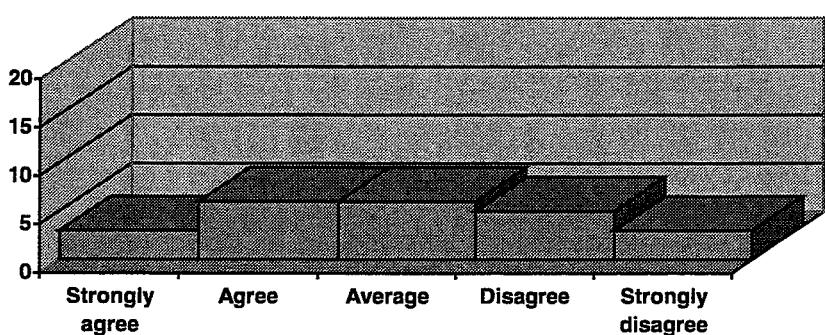
Q 22 The rationale for the 5 phases was explained.

This was apparently well dealt with by trainers. The question had been included, however, because anecdotal evidence and monitoring would indicate that the boundaries between the phases are not always well adhered to and that in some cases examiners choose to re-order the phases or even omit them. It is therefore essential that training include information about the need to maintain consistency when conducting an IELTS interview.

Q 23 Adequate time was allowed to 'digest' the descriptors.

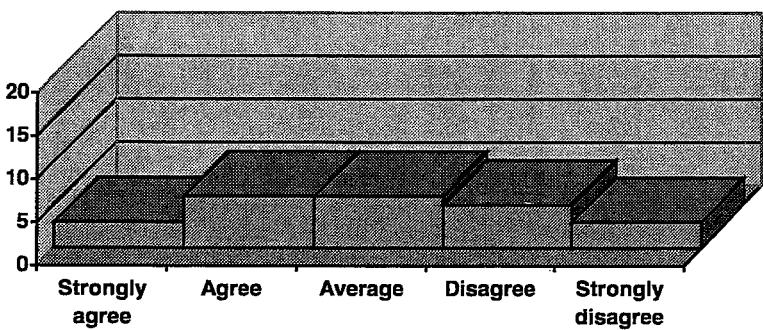
Few respondents indicated that they had required more time to take in the descriptors. This is an area of training that can be rushed, however.

Q 24 Adequate time was allowed to practise interview technique on peers.

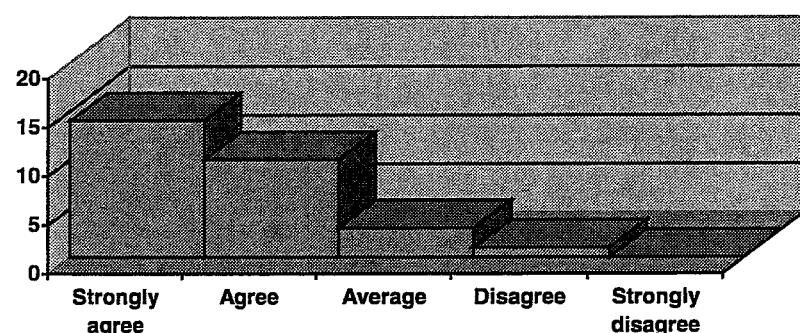


Regrettably it appears that some trainers were unable to offer sufficient time to practise the interview technique within the session. This may need to be underlined in the guidelines as the experience is invaluable and has implications for the validity of the test.

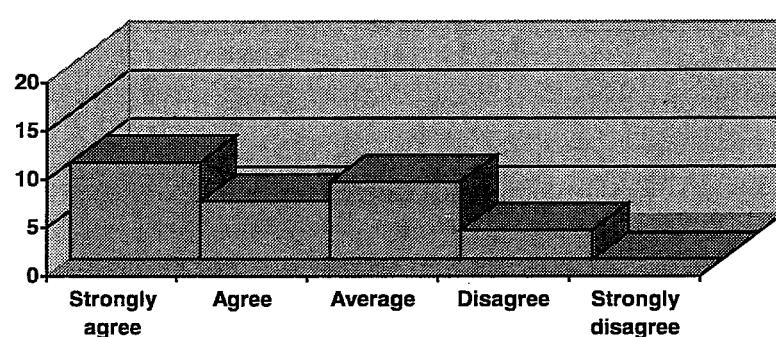
Q 25 Adequate time was allowed to prepare for the 'mock' interviews.



A couple of centres were unable to arrange for participants to undertake mock interviews, though they were urged to do a couple in their own time before working as examiners. The need to provide 'dummy' candidates cannot be over-emphasised and centres should treat this as a priority before arranging training.

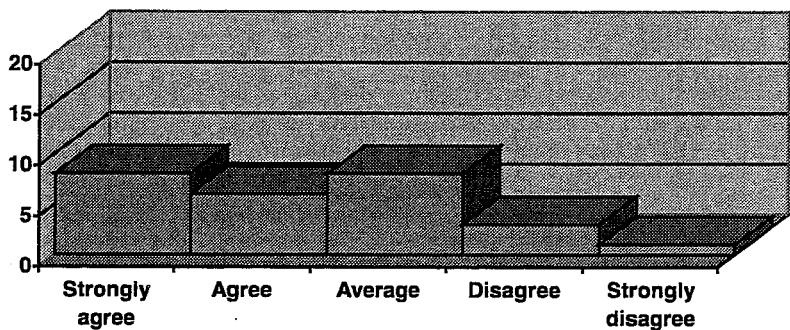
Q 26 The interview snippets (extracts from interviews) were helpful.

The question had been included because there is some controversy about whether to show parts of an interview rather than focusing on whole interviews. The benefit of using snippets is that one can focus on the individual phases and also offer a much wider range of interviews and if well handled in the training, the snippets can be very effective. Most people reported that the snippets were useful. Those who asked for more full interviews would need to be able to offer more time.

Q 27 I would have benefited from watching more low level interviews.

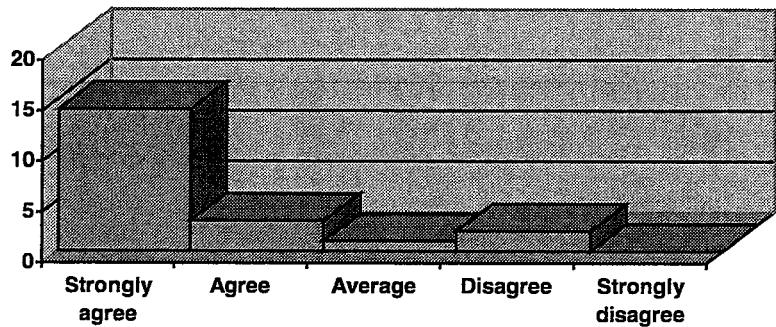
Many participants felt that more low level interviews would have been helpful. The responses tended to reflect the teaching experience and the type of candidate being met.

Q 28 I would have benefited from watching more high level interviews.



This question produced a mixed set of responses. Comments at the end of the survey indicated that more high level interviews (around Band 7) would be helpful for examiners working in India and Pakistan where many candidates are aiming for medical courses in the UK for which a Band 7 score is the prerequisite.

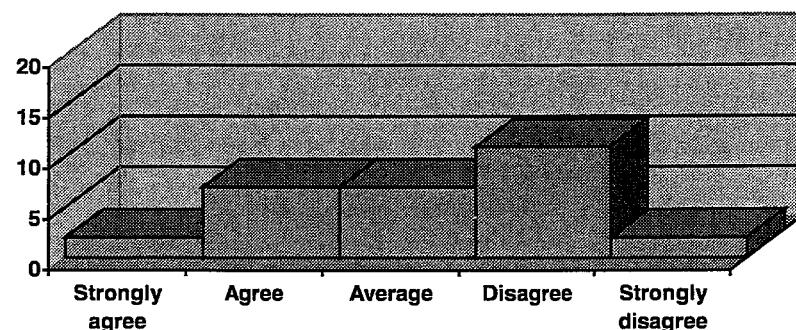
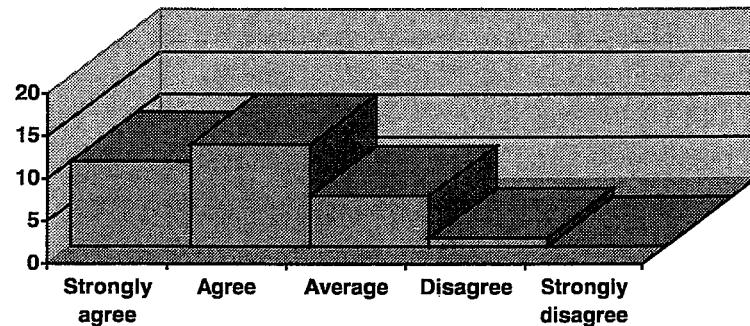
Q 29 The mock interviews on dummy candidates were helpful.



Some centres were unable to arrange for mock interviews and no response was received. The majority of participants found the experience very useful and commented on this at the end of the survey.

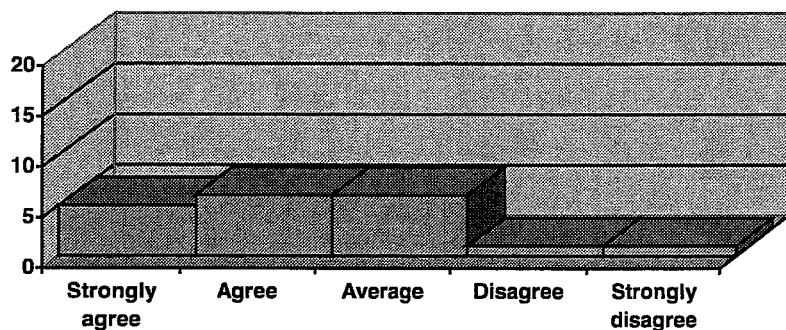
Q 30 I experienced difficulty grasping the IELTS levels.

This question was included to see if participants were generally prepared to embrace the IELTS levels as there is sometimes resistance in this area with participants feeling that they have their own idea of how the bands operate. The responses would indicate that not everybody is comfortable with the speaking bands.

**Q 31 By the end of the day I felt comfortable with the rating procedure.**

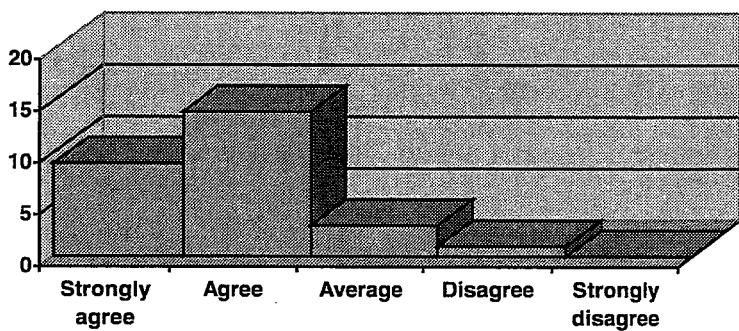
As with the response to Q.15 on the Writing procedure, not all participants felt that they were completely at ease with the procedure. Centres may need to address this problem by ensuring that refresher courses and follow up meetings take place regularly.

Q 32 I benefited from the feedback from the mock interviews.



Among those who conducted mock interviews, not all reported that the feedback was useful. It is the one opportunity that participants get to compare real ratings if they have interviewed the same candidate as someone else in the group. Aspects of inter-rater reliability can then be discussed. Perhaps more time needs to be built into the program to discuss the experience because Q.29 would indicate that they all found the mock interviews useful.

Q 33 I felt adequately prepared to undertake the certification tapes.



Most responded positively to this question though one or two people felt that the training had not addressed all the certification issues.

Q 34 Which interviews did you find the most useful? Why?

- *Mock interviews*
- *Asian interviews on video*
- *Snippets*
- *Discussing ratings with trainer*

Q 35 Which aspects of the training were most helpful in guiding you towards accurate assessment of speaking proficiency?

- *Mock interviews very useful.*
- *Watching and rating videos.*
- *Doing the certification tapes.*

General comments:

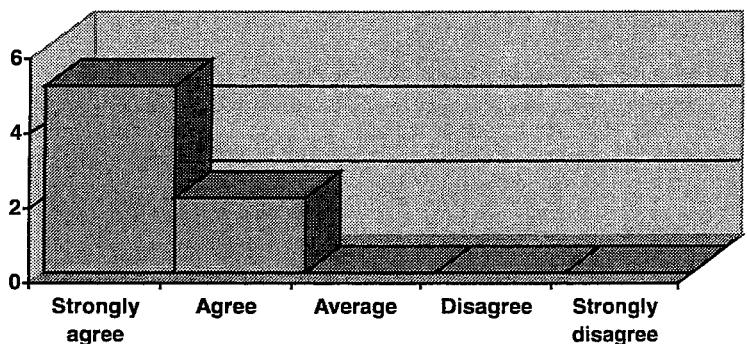
- *It is difficult to go from watching the video interviews to rating the certification tapes. We should use some audio tapes in the training session.*
- *More direction is required on what kind of questions to ask.*
- *It is very useful to compare the European candidates with the Asian candidates.*
- *I found the descriptors rather vague.*
- *One day is far too rushed. We needed two days.*
- *The mock interviews were very useful.*
- *The homework was very good.*
- *The certification tapes are of poor quality and this makes it hard to do them well.*

5.2 The Post Training Trainer Questionnaire

Respondents were asked to rate the following propositions from 1 to 5 with 1 being '*strongly agree*' and 5 being '*strongly disagree*'. In some cases, no response was provided.

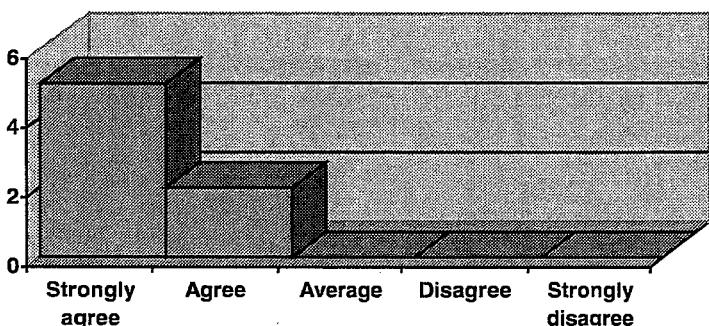
General Comments

Q 1 The two day format is satisfactory.

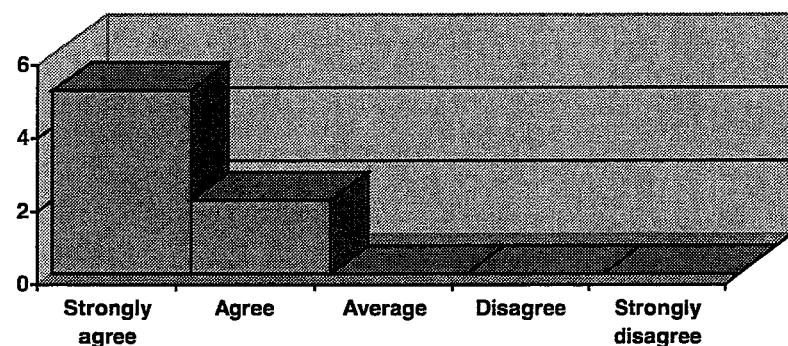


Most trainers like the format.

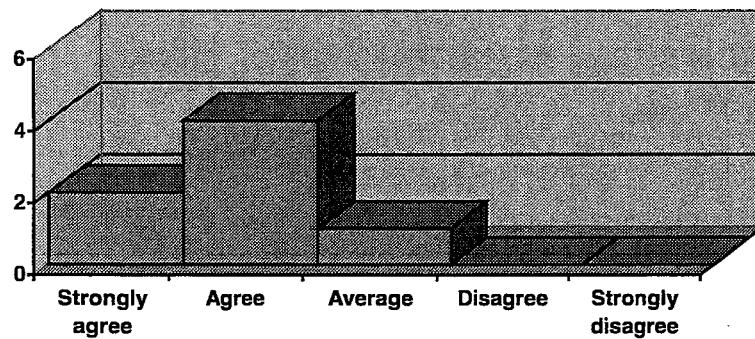
Q 2 All participants were able to participate fully in the activities.



Q 3 The accommodation for the training was satisfactory.

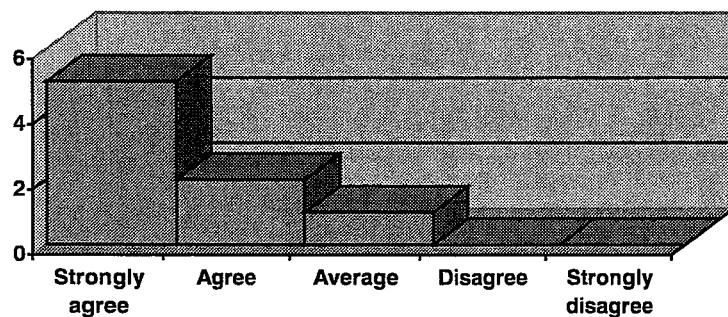


Q 4 There were no problems arranging the 'mock' interviews.



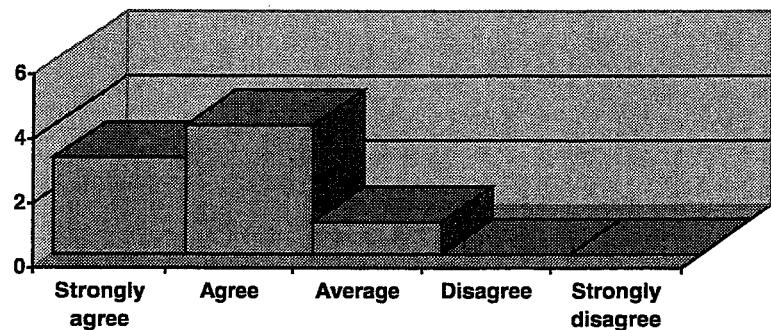
Arranging the mock interviews does impose some difficulties on trainers.

Q 5 The General layout of the training materials is clear.



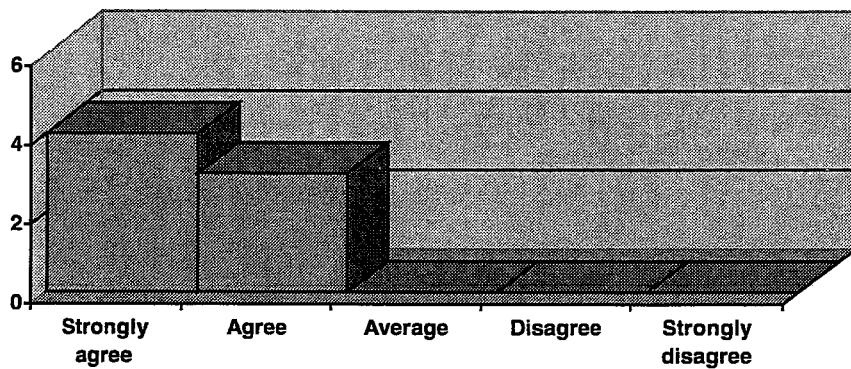
Most trainers are happy with the layout.

Q 6 The sample timetable is manageable.

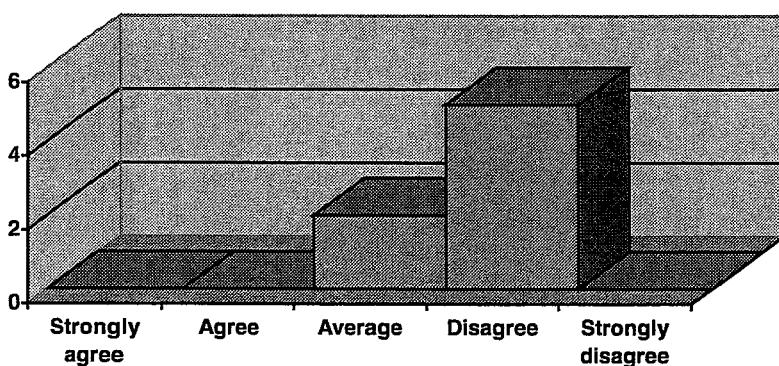


Some trainers felt that less time should be given to the Writing day.

Q 7 The materials offer adequate guidance.



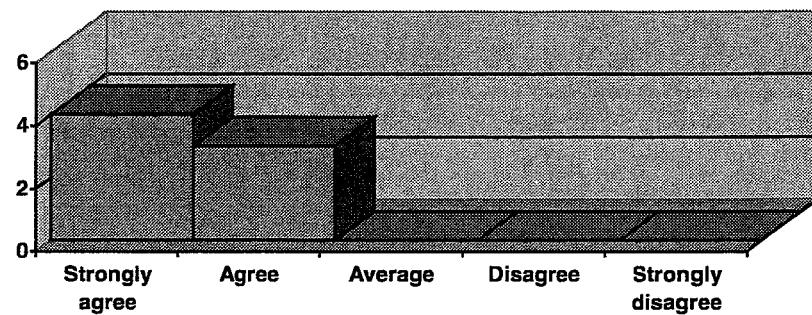
Q 8 The materials are too prescriptive



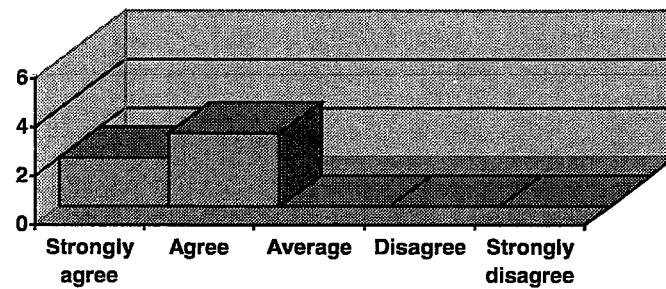
Most trainers did not find them overly prescriptive. One commented that there was considerable opportunity for the trainers to interpret the guidelines and bring their own ideas and training style to the sessions. It was intended that the guidelines be adhered to fairly rigidly to ensure standardisation of delivery.

Day One – The Writing Module

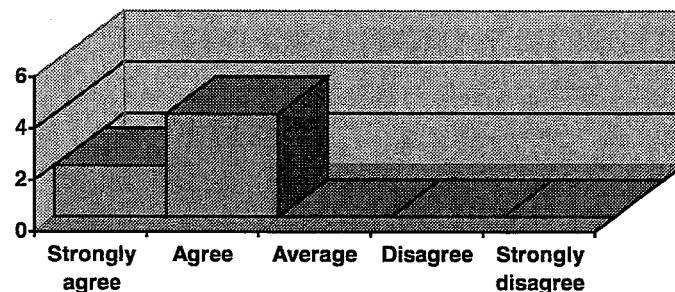
Q 9 The scripts provided for Academic Module Task 1 were satisfactory.



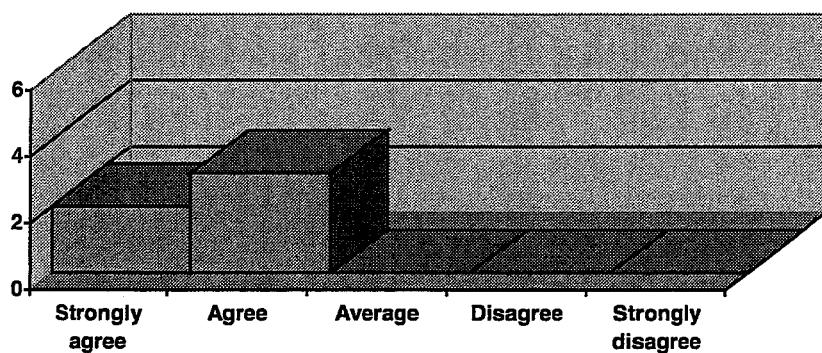
Q 10 The scripts provided for Academic Module Task 2 were satisfactory.



Q 11 The scripts provided for General Training Task 1 were satisfactory.

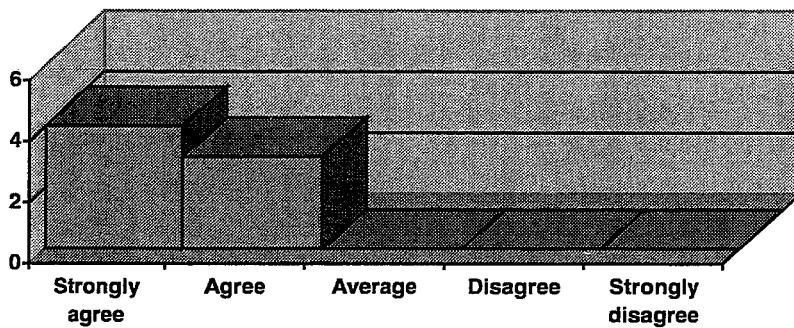


Q 12 The scripts provided for General Training Task 2 were satisfactory.

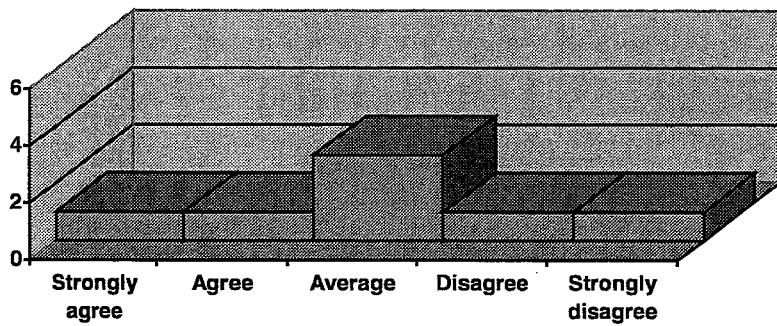


Not all trainers responded. Comments would indicate that General Training scripts were the most difficult to deal with in training.

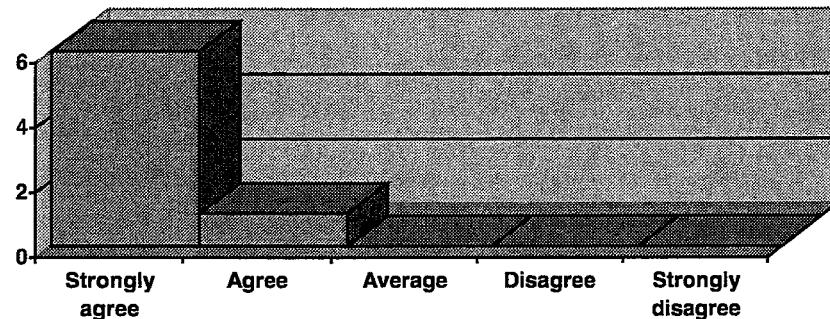
Q 13 The participants were receptive to the procedure (ie moving from group work to individual rating by the end of the day.)



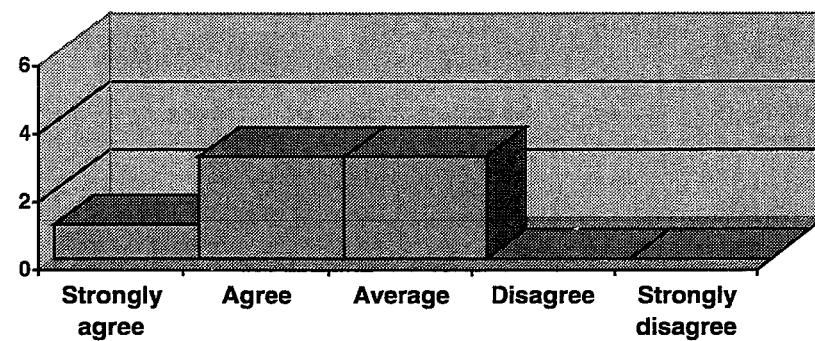
Q 14 I would have preferred more work on 'problem scripts'



Q 15 Most people completed the 'homework' supplementary scripts.



Q 16 I felt confident that the participants were rating accurately.



One trainer commented that not all those who appeared to be on target in the training session were automatically successful in certification.

Q 17 Were there any scripts which you felt did not 'work'? If so, which ones.

Most trainers did not think so, but one offered this information.

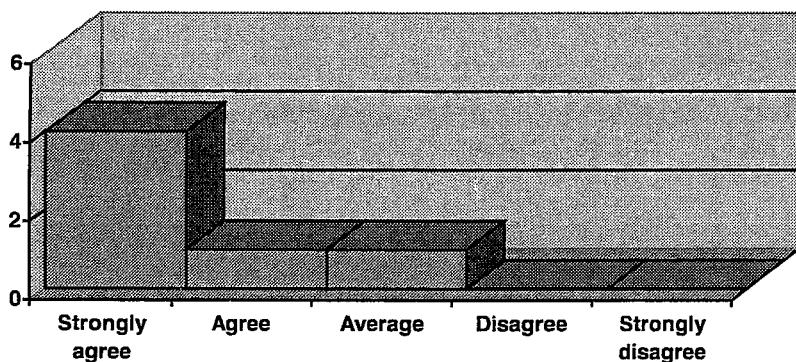
- *Academic Writing Task 1 Script 6*
- *Academic Writing Task 2 Script 6*
- *General Training Task 1 Script 2 and Script 5.*
- *General Training Task 2 Script 4.*

Q 18 Which aspects, if any, of the writing session did the participants find difficult?

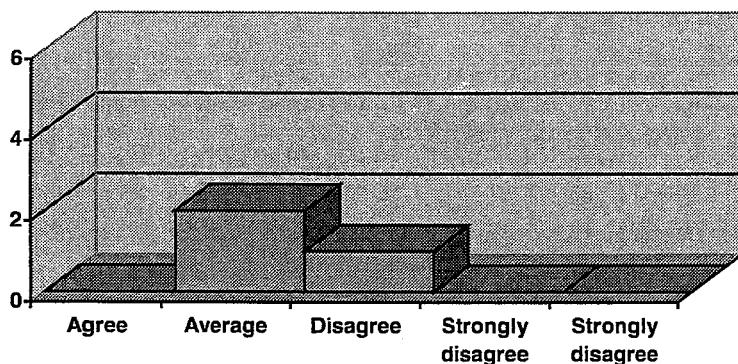
- *Switching from Academic to General Training*
- *Penalising underlength scripts if the writing is otherwise good*
- *Dealing with underlength scripts.*
- *Rating General Training Task 1.*
- *Slogging through it! Not always agreeing with the ratings and comments given.*

Day Two – The Speaking Module

Q 19 The introductory interview (Band 6 Thai candidate) is a useful starting point.

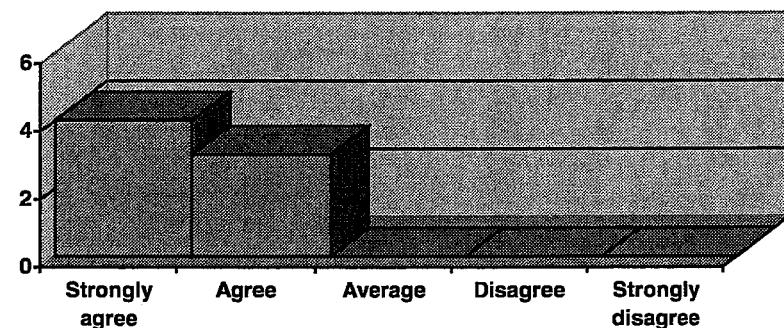


Q 20 I like to give the participants the official rating of the first interview to set the level.

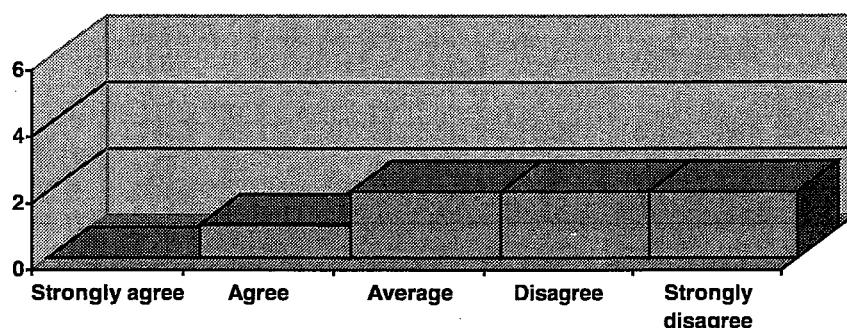


This was the recommended practice, so it is interesting to see that some trainers did not use this approach.

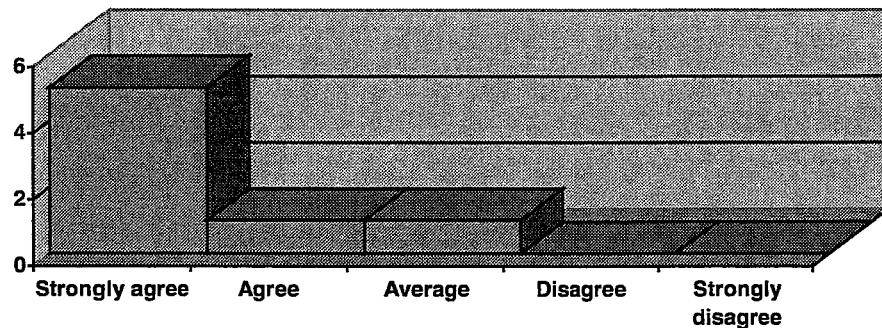
Q 21 The use of snippets (each phase) rather than full interviews is appropriate.



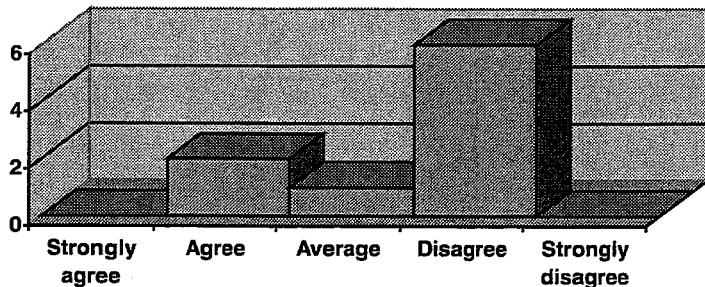
Q 22 The mock interviews were difficult to set up.



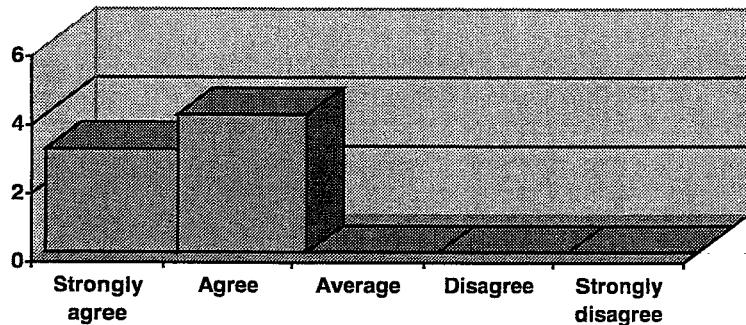
Q 23 The mock interviews proved very helpful.



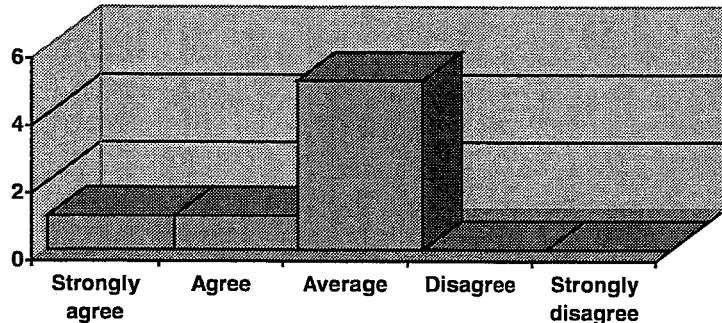
Q 24 The participants had difficulty interpreting the descriptors.



Q 25 I felt confident that the participants had grasped the interview format (the phases).

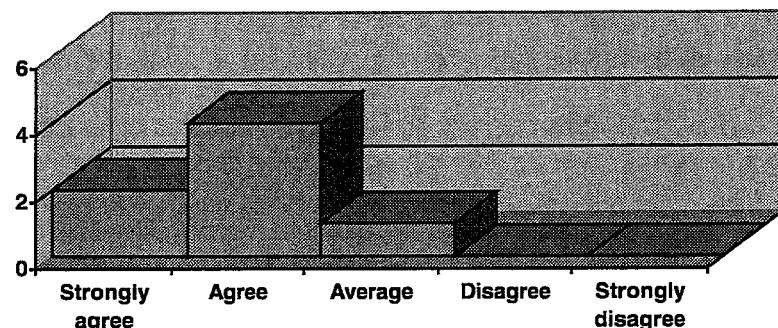


Q 26 The participants were able to bring candidates up to their linguistic ceiling.



Examiners were not confident that participants were all capable of this. Again this response points to the need to undertake mock interviews and/or peer practice interviews to ensure that the interview format is adhered to.

Q 27 By the end of the session the participants were rating accurately.



Q 28 Which interviews were most useful? Why?

- *The band 5 and 6 interviews as this is the critical level.*
- *Comparing Vietnamese and Indonesian Band 4 candidates.*
- *The Brazilian – lethargic candidate useful because she was such hard work!*
- *The Korean girl because she was so agreeable and it was easy to over-rate her performance.*
- *All interviews are useful. Would like more examples of high band candidates in phase 4.*
- *I would like another Band 7 as we interview a lot of doctors who require this score.*

Q 29 Did you disagree with any interview ratings? If so, which ones?

- *Serbian female (No. 3/8) - Band 6 not Band 7.*
- *Indonesian female – seems higher than Band 4. Participants often feel this is tough.*
- *Initially I disagreed with the ratings but after repeated viewings I've found the ratings more comprehensive/comprehensible.*
- *No – I didn't disagree with any.*

Q 30 Further comments:

- *An interview which is a clear example of Band 7 would be helpful.*
- *Participants requested more feedback on their interview performance.*
- *Generally found participants found rating the speaking easier than the writing. ie more agreement.*
- *Mock interviews were extremely useful – high degree of rater agreement.*
- *Outcomes of certification are very difficult to predict.*
- *Writing scripts work less well for refresher training. Need more scripts at higher levels.*

6.0 Overall Findings and Conclusions

6.1 The Participants' Survey

Overall the findings of the survey are very pleasing and indicate a good level of commitment from trainers and participants alike. All comments received from the participants, with one exception, indicated a high degree of satisfaction with the way in which the sessions were conducted and most were glowing about the trainers. Some adverse comments about the materials may be a reflection of the difficulties experienced in rating generally and may point to the need to re-introduce the concept of "problem scripts and interviews" or at least more thoroughly prescribed discussion of certain aspects of examining. An updated version of the training package could include:

- More short scripts
- More supplementary scripts for homework
- More emphasis to be placed on the need to bring candidates to their linguistic ceiling in the interview in order to be able to rate effectively (ie more rigour).
- Amended Guidelines to give direction in these areas.

Generally participants found the wisdom imparted by the trainers to be a most helpful aspect of the training, since as experienced raters they were able to give hints and tips on how to tune into IELTS.

6.2 The Trainers' Survey

The findings of the survey are useful and highlight some problem areas such as agreement on all "official" scores. It was noted that the materials are not really suitable for refresher courses, (for which they were not intended) and that there is a need for such courses and hence for dedicated material. Most trainers have found the materials relatively easy to use and certainly a great improvement on what was previously available. While not all trainers agreed with the actual timetable, they felt that having a prescribed schedule was helpful. One trainer commented that there was still room for an individual approach. The trainer training sessions were seen as a useful ingredient.

6.3 The Questionnaires

Some of the questions may not have produced the information being sought but generally they worked well and seemed to be easy and quick to complete. It is recommended that a slightly amended version of the questionnaires be included in the training package.

The questionnaires have shed light on a number of issues:

- the need to ensure that two days can be dedicated to the sessions
- the need to ensure that dummy candidates are available for the speaking session to allow mock interviews to take place
- the need to review one or two of the scripts and 'official' scores
- the need to include more supplementary scripts

7.0 Further action

- A slightly amended version of the questionnaires will be produced in early July 1998 and forwarded to IELTS Australia for inclusion in the training package.
- Data gathered from future questionnaires will be analysed as an ongoing part of this research and correlations with certification scores made.
- Additional supplementary scripts and possibly videotaped interviews need to be gathered to add to the training package. A.R.T.S. will seek funding from IELTS Australia to continue this work if it is deemed necessary as all material would need to be marked by senior examiners before it can be included.
- Additional scripts and videotaped interviews are required for refresher courses as this training package is not suitable for this purpose. A.R.T.S. will seek funding from IELTS Australia to continue this work if it is deemed necessary.