Qualifications and Curriculum Authority

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Comparability study of assessment practice

Personal licence holder qualifications

October 2006

QCA/06/2709

Qualifications and Curriculum Authority. (2006). *Comparability study of assessment practice: Personal licence holder qualifications.* London: Qualifications and Curriculum Authority.

Contents

Executive summary

Introduction

This comparability study is part of the regulatory authorities' programme of quality assurance monitoring of vocational qualifications. The study examined the consistency of assessment practices associated with the National Certificate for Personal Licence Holders. These qualifications are an essential component of the personal licence. Under the terms of the Licensing Act 2003, the supply of alcohol under a premises licence must be made or authorised by a person who holds a personal licence.

The awarding bodies offering this qualification are:

BIIAB
Education Development International plc (EDI/GOAL)
Graded Qualifications Alliance (GQAL)

Methodology

The comparability study commenced in November 2005 and concluded in March 2006. A team of three scrutineers, including one team leader, each having expertise in the assessment of vocational qualifications and appropriate experience, was recruited to examine assessment practices across college, employer and training provider centres approved to offer the qualification. Candidates receive a short but intensive course of instruction before taking the multiple-choice examination and the scrutineers were, therefore, asked to comment on the instruction methods used in order to provide feedback for the awarding bodies.

The team observed instruction methods and invigilation procedures and interviewed candidates and invigilators in 25 centres. Data collection was based on a common instrument provided by QCA.

In addition, a fourth scrutineer was recruited to carry out a comparative evaluation of the multiple-choice assessment materials used by each awarding body. Four examination papers and methods of presentation were examined and compared in a series of ways, both objective and subjective. In addition, an empirical exercise was carried out to estimate the relative difficulty of passing each test.

A representative of each awarding body was interviewed by telephone about the procedures in place for development and monitoring the tests.

Centre visit findings

A judgement was made as to whether the awarding body invigilation requirements were being met by the centres visited. The team of scrutineers judged that the overall process was acceptable in 76 per cent of the centres visited and, therefore, not acceptable in 24 per cent of the centres. The research that informs the judgements will be examined in this report.

The participating centres had prior notice of the scrutineers' visits. As a result, the scrutineers anticipated full compliance with the awarding bodies' requirements. Scrutineers recorded what was observed on the day. It was clear that, in some cases, extra effort had been made by the centres to prepare for the visit but, even then, the standards observed fell short of awarding body requirements.

Strengths

The following were identified as strengths in some of the centres visited:

- many tutors had a very good knowledge of the subject and were often experienced licensees
- high quality learning material issued to the candidates
- good adherence to security and invigilation procedures (however, see associated weakness)
- adequate and in many cases very good facilities used
- good instruction/learning methods (however, see associated weakness).

Weaknesses

The following were identified as weaknesses:

- no recent visit by the awarding body representative
- limited variety of instruction methods used (see also the associated strength)
- poor invigilation procedures applied
- variable level of help given to candidates with additional needs
- content that was not considered to be relevant to the company sponsoring the candidates was excluded, or sometimes additional content was included.

Test analysis findings

Analyses of the tests and items looked at the following features:

- content validity; item quality and its effects on test difficulty; empirical investigation of perceived item and test difficulty; the impact of cognitive and other demands on candidates.

By interview, representatives of the awarding bodies were asked about:

- question writing; test production; pass marks and item/test statistics; plans for future development.

Strengths

The following strengths were identified:

- good content validity
- good test format
- a commitment to developing item banking to improve quality and the consistency of standards.

Weaknesses

Several weaknesses were noted:

- dubious validity of some items
- frequent language errors in some tests
- inadequate writer qualification/training in one awarding body.

Strengths and weaknesses

In some areas, the assessment process showed both strengths and weaknesses:

- there was an awareness of the need to control item language, but tests presented some serious reading difficulties for candidates
- a difference in pass standards was seen, though the awarding body at risk had already identified this weakness
- there was a variation in the cognitive demands set by questions – good in one case, too low in another
- inconsistency in the use of post-test statistics to assure quality in future use of the items.

Conclusions

In the course of this scrutiny, various issues became apparent. Candidate preparation is weak and inconsistent, as is quality assurance of assessment delivery. While assessment delivered immediately after learning may be convenient, this study raises doubts about its effectiveness in relation to knowledge retention and reliability.

Adherence to awarding body guidance, appropriate course presentation and suitable assessment practice, was judged by the scrutineers to be effective in 76 per cent of the centres visited. While some good practice was observed in the quality of learning there were particular concerns raised regarding:

- invigilation and security of examination papers
- confusion over resources and support arrangements for specific types of candidates
- and course delivery that does not always meet the learning specification.

The compliance rate indicates that the majority of centres are implementing assessment to the required standards but that, overall, assessment is inconsistent.

Action taken to address the shortfalls identified would do much to improve the quality of the candidates' experience in relation to the multiple-choice tests. Although they differ slightly in how they do so, all of the awarding bodies ensure that their tests adequately cover the test specification.

There are some concerns about the nature of several types of questions used. Some seem to place excessive demand on memory. This is of particular concern when testing immediately follows teaching, so that only short-term memory is required. Other questions address information that it is unimportant, or that it is unreasonable to expect every candidate for a Personal Licence to know.

Several of the faults commonly described in textbooks and training materials for multiple-choice item writers were common in these tests. Procedures for catching these should be tightened up.

Reading difficulty was judged too high in a significant number of questions, which raises questions about their validity. This would have a particular effect on candidates for whom English is a second language.

The BIIAB test makes more intense demands in a relatively short test. The EDI test is the most comprehensive, with a greater number of questions, a greater requirement to read and the

allocation of more time to complete the test than the others. The GQAL test is the most appropriately designed test for candidates who have difficulty reading English (despite having a greater number of items that are difficult to read because of faults).

This report contains the findings from the centre visits and the comparative evaluation of the multiple-choice papers, and will be made available to the Department for Culture, Media and Sport (DCMS) and all the awarding bodies that offer these qualifications.

Recommendations

There is room for improvement in the procedures of all the awarding bodies. Many of the weaknesses noted in this study could be addressed by the introduction of robust quality assurance by the awarding bodies, and particularly by increasing visits made to centres that deliver the qualifications. The awarding bodies need to:

- consider producing guidance and/or materials to stimulate diverse methods for delivering content
- rectify factual inaccuracies in course content
- ensure robust invigilation and security arrangements for assessment
- ensure that the qualification and assessment arrangements are suitable for the candidates' needs
- GQAL needs to consider raising its pass marks to represent the same standard as the others
- during the next 12 months a review of questions should be carried out by or including an independent content expert, to identify items that address inappropriate content or appropriate content in an inappropriate way and to suggest better item types to replace them. It might also consider, in consultation with the DCMS, whether all of the topics/sub-topics in the DCMS specification deserve equal weighting. The review should consider whether it is acceptable for candidates to be tested on knowledge on the same day that they acquire it.

- The GQAL test is the least demanding in that too many of its items test only recall of facts and vocabulary. GQAL should increase its use of items that test understanding. EDI should also adjust the balance of their items slightly in the same direction.

- The awarding bodies should be encouraged to develop adequate item banking systems as quickly as possible in order to ensure the standards are maintained.

Each awarding body is asked to respond in writing to the report, indicating how it intends to address any issues of concern highlighted by the study.

PART ONE

1. Detailed centre visit findings

1.1 Introduction

Under the terms of the Licensing Act 2003, the supply of alcohol under a premises licence must be made or authorised by a person who holds a personal licence. A personal licence is granted by a licensing authority to an individual if that individual:

- is aged 18 or over
- possesses a **licensing qualification**
- has not forfeited a personal licence during the previous five years
- has not been convicted of any relevant offence or any foreign offence.

All personal licence qualifications have to be accredited by the Secretary of State for Culture, Media and Sport, and be awarded by similarly accredited bodies. The Secretary of State will only accredit qualifications that have been accredited by the Qualifications and Curriculum Authority (QCA) or the Department for Education, Lifelong Learning and Skills (DELLS).

As these qualifications are an essential component of the personal licence, the regulatory authorities have undertaken a study to compare the assessment requirements of each qualification and the assessment arrangements in centres. Assessment for each qualification is by multiple-choice examination with a pass/fail result. The structure of the multiple-choice papers and the assessment requirements do, however, differ among the three awarding bodies. Delivery in centres requires 10 guided learning hours for the mandatory component of each qualification. During this study, significant differences in assessment practice were noted in order to identify good practice, to ensure consistent demands are made on candidates, and that appropriate assessment arrangements are made in centres.

The National Certificate for Personal Licence Holders was first accredited to the National Qualifications Framework on 1 February 2005. It is currently offered by three awarding bodies: BIIAB; Education Development International plc (EDI/GOAL) and the Graded Qualifications Alliance (GQAL).

Purpose

The scrutineer team focused the study on three main areas. Its purpose was:

- to report on assessment practice for personal licence holder qualifications accredited by the Department for Culture, Media and Sport (DCMS) for licensing purposes
- to note significant differences in assessment practice where found and to identify good practice
- to make recommendations for improvements in assessment practice where the outcomes of the study suggest these are necessary.

This comparability study is part of the regulatory authorities' programme of quality assurance monitoring of vocational qualifications. As stated in *The statutory regulation of external qualifications,* the outcomes of monitoring activity are publicly reported.

Due to the time constraints and the size of the sample the outcomes of this study should be taken as indicative. Nevertheless, the findings indicate aspects of delivery requiring attention and, where appropriate notable examples of good practice that should be encouraged.

For each of the personal licence holder qualifications accredited by the Department for Culture, Media and Sport, the team of scrutineers reviewed the following:

Assessment

- assessment requirements
    - structure of multiple-choice tests
    - coverage of assessment criteria
    - compulsory questions
    - mark scheme
    - demand on candidates
- maintenance of awarding body question bank
- mechanism for secure delivery, collection and marking of test papers
- mechanism for issue of results and delivery of certificates
- centre satisfaction with results turnaround time.

Centre operations

- 10 guided learning hours
- examination rooms – location and physical environment
- tests conducted fairly – controls in place
- candidates' identities checked
- security of test papers
- procedures for making reasonable adjustments and accessing special consideration.

Scope

The study covered the Level 2 National Certificate for Personal Licence Holders accredited by the DCMS and offered by BIIAB, EDI and GQAL. Centres were selected at random across England and Wales on the basis of information supplied by the awarding bodies.

1.2 Visit methodology

The team of scrutineers visited 25 approved centres across England and Wales (two in Wales), between November 2005 and March 2006.

At each approved centre the scrutineers examined the instruction methods, assessment security and invigilation practices applied. They interviewed an average of five candidates and the key staff associated with the course in each centre.

The scrutineers were required to make judgements as to whether centres were adhering to the awarding body guidelines and requirements, and to record their judgements on the data collection instrument. In particular, examination paper security and invigilation arrangements were observed and judgements recorded.

1.3 Centre sample profile

The original intention was to construct a representative sample of centres to visit, matched against the number of centres approved by each awarding body and spread across the different types of assessment centre. The centres chosen were to be selected at random from the information supplied by the awarding bodies. Centres can be registered with a number of awarding bodies to deliver their qualifications.

The actual make-up of the final sample was dictated by the fact that many centres were not offering the courses when contact was made. Overall, 38 personal licence holder centres were contacted and 25 visits made.

A further difficulty was that many of the centres in the sample were registered with more than one awarding body. Additional time was required to confirm that a centre was offering the personal licence holder qualification for the specified awarding body.

The following table outlines the centre selection procedure and the final numbers involved.

Table 1. Approved centre selection for study sample

Centre selection process	Number of centres
Original selection from lists supplied by awarding bodies	41
Centres that were not currently delivering the qualification	9
Centres not to be visited (visited recently by QCA on other matters)	3
Course cancelled	4
Centres visited	**25**

Centre type

The following table shows the types and numbers of approved assessment centres participating in the study.

Table 2. Approved centre type profile

Centre type	Number of centres
Training provider	13
Employer	10
College	2

Awarding body

The following table shows the number of centres visited for each awarding body, and reflects the proportion of centres each awarding body had approved at the commencement of the study. Some centres deliver their courses and assessments through satellite sites, and Table 3 should not therefore be taken to indicate awarding body market share.

Table 3. Approved centre profile by awarding body

Awarding body	Number of centres visited
BIIAB	18
EDI	1
GQAL	6

1.4 Programme delivery

The results of the observations made by the scrutineers in relation to the programme delivery are recorded in this section. At all times the scrutineers made judgements based on the guidance supplied to centres by the awarding bodies. In most cases the results are expressed as a percentage of the number of awarding body requirements fulfilled by the centre. A result of 100 per cent would, therefore, imply the centres were judged to be fully complying with the guidance given.

Scrutineers' judgements were aggregated to produce the results for overall compliance in each table; these figures cannot be calculated by taking the mean of the results for each of the three awarding bodies.

Deliverer

Scrutineers collected information on whether the courses and assessments were being sub-contracted out. In all the centres visited, the registered centre carried out the training. In many centres this involved 'buying in' the expertise, but none of the training courses was directly sub-contracted to another company.

Mode

Information was collected relating to the mode of course delivery. Most personal licence holder courses were run on one day with the examination taking place at 4.00pm. Most centres sent out the course books a week or more in advance, but some did not. Courses were presented on weekdays and at weekends.

One centre spent no time teaching and candidates appeared only to take the examination (BIIAB).

One college centre ran the course over one and a half days, with the two sessions scheduled over two weeks (BIIAB). The final half-day was used for revision and the examination. The candidates interviewed particularly appreciated this approach.

Table 4. Mode of delivery

Awarding body	Centres providing training %
BIIAB	94
EDI	100
GQAL	100
Overall	**96**

Guided learning hours

The Learning and Skills Council define guided learning hours as:

'... all times when a member of staff is present to give specific guidance towards the learning aim being studied on a programme. This definition includes lectures, tutorials and supervised study in, for example, open learning centres and learning workshops. It also includes time spent by staff assessing a learner's achievements, for example in the assessment of competence for National Vocational Qualifications (NVQs). It does not include time spent by staff in the day-to-day marking of assignments or homework where the learner is not present. It does not include hours where supervision or assistance is of a general nature and is not specific to the study of the learners.'

Funding Guidance for Further Education in 2006/07

Personal licence holder courses are recommended to offer 10 hours of guided learning time. Only one of the centres visited met this requirement (BIIAB). Almost all of the others ran the course on one day from 9.00am to 5.00pm, with the examination taking place at 4.00pm. After subtracting time for breaks and administration, the actual time spent providing guided learning was between 5 and 6 hours. However, in most cases the scrutineers judged this to be acceptable where the books had been sent out in advance to candidates along with guidance. Two centres (one BIIAB and one GQAL) did not send notes in advance.

Table 5. Guided learning hours

Awarding body	Centres offering recommended learning time %
BIIAB	83
EDI	100*
GQAL	83
Overall	**84**

** Based on one centre only*

Location

The scrutineers collected data to record where the programme was delivered. The percentages in Table 6 relate to the number of courses presented at the registered centre, as opposed to other locations, such as rented rooms or hotels.

Table 6. Course delivered at registered centre

Awarding body	Course delivered at registered centre %
BIIAB	72
EDI	0 *
GQAL	17
Overall registered centres	**56**

** Based on one centre only*

Facilities

The facilities were assessed to ensure they were adequate for running training courses. Overall, a very good standard was observed. One BIIAB centre was found unacceptable because it was not heated.

Table 7. Facilities used for the training course

Awarding body	Centres judged to be providing adequate facilities %
BIIAB	94
EDI	100*
GQAL	100
Overall	**96**

** Based on one centre only*

Delivery methods

Because of the intense nature of the short course leading to the examination, it was deemed appropriate that the scrutineers should observe and comment on the delivery methods used and feed this back to the awarding bodies. The delivery of learning was considered from the points of view of content and style.

The content presented during the vast majority of the courses was substantial. Comprehensive books and slides prepared by the awarding bodies were used in many cases. Almost without exception, the scrutineers commented favourably on the depth of knowledge and experience of the tutors. One less favourable aspect however, was that many centres took back the books they had issued once the course was complete. Prior to the courses, many centres encouraged self-study by the candidates.

The style of the delivery in many centres was equally encouraging, with group activities and interaction playing an important part. However, centres made excessive use of presentations delivered via overhead projectors (OHPs). For such intensive courses this may appear to be the only way to cover the content. However, some centres broke up the presentations with activities, such as quizzes, to very good effect.

One example of good practice was seen at a GQAL centre. The tutor had prepared thorough OHP slides relating to the course and made good use of a flip chart. A good Q&A technique was used, all candidates were involved in discussions and a mock examination was held to very good effect in preparing the candidates.

Table 8. Delivery methods

Awarding body	Delivery methods judged to be appropriate %
BIIAB	89
EDI	100*
GQAL	100
Overall	**96**

** Based on one centre only*

1.5 Invigilation and security guidance from awarding bodies

There were a number of differences in the guidance supplied to centres by the awarding bodies with respect to the invigilation and security procedures. The following table highlights these differences. The scrutineers based their decisions on the guidance given.

Table 9. Comparability of guidance

Awarding body requirements	BIIAB	GQAL	EDI
Who can invigilate	Should not be tutor	Not tutor or teacher	Not trainer
Should invigilators be trained	'Suitably qualified and experienced staff'	Principal invigilator must attend *Training the Invigilator* course.	Suitably qualified and experienced staff
Paper delivery methods	Sealed envelope by post.	Sealed envelope by post.	Sealed envelope by post
Security arrangements	Opened in front of candidates. On examination completion Q&A sheets to be sealed in envelope within examination room	Opened in front of candidates. Question papers to be resealed before answer sheets faxed for marking	Opened in front of candidates. Question papers sealed and returned
Paper return methods	By post within one working day.	Fax answer sheets. Ensure security during process	Scanned and emailed
Provision for additional needs candidates (Learning difficulties)	25% extra time. Reader. Questions on audio tape. An amanuensis (writer)	25% extra time. Reader. Tape. Writer. Reader must not be tutor and the writer must not be a relative	25% extra time. Reader. Scribe. Use of a keyboard
Provision for additional needs candidates (ESL)	Additional time reader. English or bilingual dictionary (non- electronic)	Translators or interpreters not allowed. Reader (in English), no extra time. Bi-lingual dictionaries (non-electronic)	Non-electronic translation dictionaries. Extra 25% of time if candidate in UK for less than 2 years
Procedures for readers and for writers	Separate room. If reader is also writer, extra invigilator required (not tutor, not relative)	Reader/writer should not act as invigilator. Separate room. Extra invigilator	Separate room. Extra invigilator, not the tutor or relative of the candidate

1.6 Assessment security and invigilation

Pre-assessment security

The scrutineers observed:

- the procedures used for delivery of assessment materials to the assessment location
- the security arrangements used when the materials were opened/made available to invigilators/staff
- the time when the materials were opened/made available to invigilators/staff.

Papers from BIIAB and GQAL were delivered in sealed envelopes and in all the centres visited were kept secure until they were opened in front of the candidates at the start of the examination. This appeared to be a routine method that worked very well.

The EDI centre had an equally robust system in which the papers were supplied in 'tagged' bags from the centre HQ, and delivered to the training and assessment location. The bags, locked in a car boot, were opened in front of the group at the appropriate time.

Table 10. Pre-assessment security 100%

Awarding body	Centre compliance with AB guidance %
BIIAB	100
EDI	100
GQAL	100
Overall compliance	**100**

Invigilation

Invigilation procedures were a key part of the observations made by the scrutineers. The scrutineers recorded details on the level of centres' compliance with the specific awarding body requirements. These related to the process in general, but in particular:

- candidate identification (ID)
- number of invigilators, whether they were the course teachers, and any others present
- time-keeping arrangements
- information supplied to the candidates before start of examination (ie fire, food, toilet, illness)
- arrangement for removing bags and notes from desks
- awarding body information made available to the invigilators.

In all cases the ratio of candidates to invigilators was acceptable.

At four BIIAB centres the invigilation procedures did not correctly follow the specified guidelines. At two centres the course tutor acted as invigilator, which is not recommended by the awarding body. At the same two centres the examination procedures were outlined poorly or not at all. At a third centre, a very inexperienced invigilator had to look after 15 candidates. When an issue arose relating to ID, the examination had to start late as no back-up was available. At the fourth centre, no examination procedures were outlined and no clock was used. In addition, no instruction was given about removing notes or bags from the desks. This invigilator admitted to having had received no training.

Invigilation at all the GQAL and EDI centres was judged to be meeting the guidance. In most cases it was carried out to a very high standard. A few small concerns were noted. For example, one or two centres did not specify a particular clock to be used for the timing, and procedures could have been explained more clearly. In some cases the tutor acted to guide a less experienced invigilator. In some centres where ID was checked immediately before the examination, this process took a long time. However, ID was always checked carefully.

Table 11. Invigilation

Awarding body	Centre compliance with awarding body guidance %
BIIAB	72
EDI	100*
GQAL	100
Overall compliance	80

** Based on one centre only*

Facilities used for the examination

In most cases the room used for instruction was rearranged for the examination. The scrutineers checked that the conditions were adequate (with regards to room temperature, light and noise, for example) and that the space between desks was appropriate.

One BIIAB centre was deemed below standard because it was so cold. A small fan heater was brought in five minutes before the examination, but it was inadequate for the size of room. The tables also had a spongy surface.

Table 12. Examination facilities

Awarding body	Centres judged to be providing adequate facilities %
BIIAB	94
EDI	100*
GQAL	100
Overall	**96**

** Based on one centre only*

Additional requirements

Arrangements for candidates with additional requirements varied considerably and some centres were confused over what was allowed and what was not. In most cases an additional person was made available and worked with candidates in a separate room.

In most cases noted during the visits, dyslexic candidates and those with English as a second language (ESL) were allowed extra time if requested.

At two BIIAB centres, staff did not know the procedures to follow in order to accommodate candidates with additional requirements. Both members of staff were very inexperienced.

One centre referred ESL candidates to a local college, as the centre felt the college would have facilities that would better cater to them. Most centres try to find out if extra help will be needed when the candidates apply for the course.

Table 13. Arrangements for candidates with additional requirements

Awarding body	Centres operating within awarding body guidelines %
BIIAB	83
EDI	100*
GQAL	100
Overall	**88**

** Based on one centre only*

Marking and post-assessment security

The scrutineers judged the arrangements for securing completed test papers to be in compliance with awarding body requirements.

Special envelopes are supplied by BIIAB and these are used to return the papers to the awarding body immediately after the examination or at the earliest opportunity.

EDI papers are returned in tagged and sealed bags to the centre's main office, where they are scanned and sent to the awarding body by email.

GQAL answer sheets are faxed for instant marking. Question papers are returned in sealed envelopes. Results are normally returned within the hour. If a pass is achieved, the candidate is given the result as soon as it is received. If a candidate fails, an immediate re-sit using a different examination paper is offered to the candidate. The paper is returned for marking using the same faxing procedure. The efficacy of an immediate re-sit is discussed later on in this report.

Table 14. Marked papers returned to awarding body

Awarding body	Centres following awarding body guidance %
BIIAB	100
EDI	100*
GQAL	100
Overall	**100**

** Based on one centre only*

1.7 Other issues observed

Exemptions

No exemptions are allowed for the personal licence holder course or examination.

Minor concerns

The scrutineers observed a range of issues that arose, particularly during the assessment. Centres dealt with many of these minor issues competently.

An invigilator at a BIIAB centre with only two weeks' experience spoke to the scrutineer about a candidate who did not have the correct ID, but who had offered to fax it to the centre later on. The invigilator asked the scrutineer for advice. The scrutineer suggested that the invigilator should consult with a superior at the centre. Had the scrutineer not been present, the candidate may have been accepted for the examination and could have by-passed the ID requirements. In the scrutineer's opinion, this problem arose because the invigilator had not received adequate training.

During one BIIAB examination significant noise began to issue from an adjoining room. The invigilator had no means of contacting anyone about stopping the noise without leaving the room, and tried shouting a request for quiet to no effect. Eventually, at the request of the invigilator, the scrutineer went next door and dealt with the problem. This could have presented a serious issue, if a scrutineer had not been present.

At one GQAL centre, a candidate indicated all answers on the question sheet, by ringing the letter of the chosen answer. The candidate was instructed by the invigilator to transfer the answers to the proper answer sheet. This was entered on the examination incident log and sent to the awarding body.

1.8 Awarding body results and appeals procedures

Results

The turnaround time for results and certificates was investigated.

BIIAB centres varied a little on the time required to issue results and certificates. In general, results were returned within one to three working days, and certificates were returned within one to two weeks. Centres were satisfied with this.

The EDI centre reported that EDI takes about two weeks (up to four weeks on some occasions) to return results. The centre did not consider this to be acceptable, as the results were scanned and emailed directly to the awarding body shortly after the examination.

GQAL results are returned by fax within 30 to 60 minutes of the examination. The time for return of certificates seemed to vary across centres, with centre staff reporting that this ranged from two to three days, up to as much as two to three weeks.

Table 15. Results and certificates turnaround time

Awarding body	Centres reporting acceptable turnaround time %
BIIAB	100
EDI	0*
GQAL	100
Overall	**96**

** Based on one centre only*

Appeals procedure

There was little consistency regarding how candidates were made aware of the awarding body appeals procedure. There was also some confusion among centre staff as to whether one existed. The results in the following table show the outcomes.

In many centres, candidates were told that an appeals procedure existed and that a note was pinned on the wall to provide further details if needed. This is good practice.

Table 16. Appeals procedure

Awarding body	Candidates made aware of procedure %
BIIAB	81
EDI	0*
GQAL	83
Overall	**64**

** Based on one centre only*

Candidates

The team was asked to interview a representative number of candidates (usually five or six) at each of the approved centres visited. The total number of candidates interviewed was 110. The candidates were asked to give feedback on their general experience, the course content and the assessment process. The following table shows the gender profile of candidates taking the course and examination.

Table 17. Candidate profile

Gender	Percentage of total cohort %
Male	75
Female	25

General experience

The candidates' perspectives of the course in general were very positive across all three awarding bodies.

At one BIIAB centre, the lack of heating, poor quality of tables and confusion on the day of the examination were considered to be unacceptable. This was not the norm for the centre, however.

At the one EDI centre examined, a number of candidates expressed dissatisfaction with the cost of the course and arrangements, in general. However, they all considered the course content to be very good and well presented.

One candidate at a GQAL centre was an experienced licensee, who had missed out on claiming 'Grandfather Rights'. The candidate claimed to be very pleased with the course.

Table 18. Candidates' general experience

Awarding body	Candidate satisfaction %
BIIAB	94
EDI	0*
GQAL	100
Overall satisfaction	**92**

** Based on one centre only*

Course content

At one BIIAB centre, all candidates interviewed said some content was not covered (eg relating to drugs, and children on licensed premises).

A typical comment, reflecting the responses received across all awarding body centres, was '...it was more interesting than I was expecting. I have spent a long time in the trade but I still learnt a lot'. Many candidates also said that, '...it was useful to have the handbook a few days before the course, as there is a lot of information to absorb in one day'. On some occasions, however, candidates said that they had not received the book in advance.

Table 19. Candidates' satisfaction with course content

Awarding body	Candidate satisfaction %
BIIAB	94
EDI	100*
GQAL	100
Overall satisfaction	**96**

** Based on one centre only*

Assessment process

Candidates' perspectives on the assessment process and arrangements for invigilation were, in general, very positive.

Across all awarding bodies, several candidates commented that '... the language used in the questions was confusing.' Several centres also considered that increased emphasis on using plain English would be an improvement.

Many candidates expressed their dislike of the BIIAB 100 per cent pass mark for part A. A typical candidate's comment was: '...we fail if just one silly mistake is made, even if we get all the part B questions correct.'

Candidates at the EDI centre said they felt as if they had wasted two hours filling in forms.

Table 20. Candidates' satisfaction with the assessment process

Awarding body	Candidate satisfaction %
BIIAB	83
EDI	0*
GQAL	100
Overall satisfaction	**84**

** Based on one centre only*

2. Overall analysis

In general, although there is room for improvement in some areas, the general outcome of the scrutineers' research is quite positive. Some centres were considered less than adequate. However, many of the centres that were judged critically were making an attempt to comply with

awarding body guidance. Some further guidance, and ideally a visit from an awarding body representative, would solve many of the problems that were noted by the scrutineers.

2.1 Overall results by awarding body and centre type

Awarding body

The following table shows the percentage of centres at which the scrutineers judged that the awarding body guidelines were being followed.

The results indicate that some BIIAB centres have issues to resolve. These are highlighted in the main body of the report and in the Recommendations section.

Table 21. Centres following awarding body guidelines

Awarding body	Following guidelines %
BIIAB	67
EDI	100*
GQAL	100
Overall	**76**

** Based on one centre only*

Centre type

The following table compares the level of centres' compliance with their awarding body requirements by centre type. The variation in centre type numbers in the sample must be considered when comparing these agreement rates. The results show that training providers appear to be doing a better job than the colleges and employers. The results relating to colleges were affected by the small number of centres visited.

Table 22. Centres following awarding body guidelines by centre type

Centre type	Centres following guidelines (no. of centres) %
Training provider	92 (13)
Employer	56 (9)
College	67 (3)
Overall	**76**

2.2 Strengths

The following were identified as strengths in some of the centres visited.

Many tutors had very good knowledge of the subject and were often experienced licensees
Most trainers were very experienced and worked hard to show the benefit of the new award to the licensed trade. Many had experience of the industry and were able to work well with the particular client group. Most had excellent credibility and gave the impression that they believed in what they were teaching and not simply delivering a course.

Good quality material issued to the candidates
Many centres had created additional high-quality handouts to support the presentations. The material supplied by the awarding bodies was very good.

Good adherence to security and invigilation procedures (however see associated weakness)
The majority of centres were doing a very good job in this respect. In some cases in which errors were made, this were a result of ignorance rather that a wish to commit fraud or cheat. Some centres had very comprehensive procedures in place to cover different awarding body requirements.

Adequate, and in many cases very good, facilities used
This is stated as a strength because it was considered potentially problematic in the industry. In many cases very high-quality, purpose-built training rooms were used.

Good teaching/learning methods (however, see associated weakness)
In some centres short mock examinations were used (under examination conditions) to help candidates prepare. Some centres used pre-tests to check that candidates had done preparatory work. One scrutineer remarked, 'I sat in on a course for nearly five hours and enjoyed the presentation and professionalism of the show'. Many trainers used a variety of training techniques to keep learners interested, such as buzz groups and short quizzes.

2.3 Weaknesses

The scrutineers also observed a number of weaknesses. Attention to these would improve the quality of assessment and overall provision.

No recent visit made by the awarding body representative

Some centres had never had contact with a representative from the awarding body. Many said they would welcome advice and the chance to ask questions. The scrutineers were told, on several occasions, that they were the first observers to visit the centre. Many of the problems raised in this report could be dealt with if visits were made by representatives of awarding bodies.

Limited variety of teaching methods used (see also the associated strength)

Many of the candidates did not have the frame of reference necessary to comment on the teaching methods used. However, there was room for improvement in some centres. Some candidates (and the scrutineers) commented on over-use of OHP presentations. Notes and/or books issued to candidates were taken back at the end of the course – apparently due to the cost of providing new books to each candidate.

Poor invigilation procedures applied

Unqualified and inexperienced invigilators were among the key concerns that this study highlighted. Many had limited knowledge of the process required. In some cases it appeared that the invigilator present during the visit had been drafted in for the benefit of the scrutineer and was there simply to carry out administrative duties while the tutor ran the examination session. Common mistakes made during the assessment process included: times were not noted; no clock was used; incorrect times were announced; and tables were not arranged appropriately. In most cases these errors occurred due to ignorance of the correct procedures. In some cases, the invigilation was carried out by the course teacher (even when the awarding body specified that this should not be done). In some centres the invigilator had no back-up in the event of a problem arising (even one as basic as a candidate needing the toilet).

Variable level of help given to candidates with additional needs

There appeared to be some confusion over what extra help should be available for candidates with additional requirements, particularly those with English as a second language. However, most centres were able to cope with this aspect. A key point raised by a number of sources was the examination language and style. If this were to be improved, the incidence of ESL problems, and indeed other additional requirement issues, would be much reduced.

Content being omitted or additional content included

At a small number of centres, subjects were covered in teaching, but not in the examination. In addition, some trainers concentrated on material related to their own area of expertise (for example, retail) and de-emphasised material that did not apply to that particular sector. This is clearly wrong, since the qualification is intended to be portable and to apply to a range of sectors. Some trainers included content on the detail of licensing legislation, which it was felt went beyond

what was necessary for adequate completion of the examination or the candidates' likely working role.

Comment

It should be noted that, for every centre where a weakness was identified, a corresponding strength or at least an adequate performance could be found in another. This implies that the overall quality of provision is inconsistent and there is room for significant improvement.

3. Conclusions

Adherence to the awarding body guidance, appropriate course presentation and suitable assessment practice was judged to be effective in 76 per cent of the centres visited. A number of key strengths were identified. Nevertheless, there were also weaknesses in a number of areas.

Candidate preparation is weak in some centres, but of particular concern is the invigilation process. Awarding bodies must do more to ensure their guidance is followed.

- Delivery methods tend to be unimaginative and fairness in assessment is at risk where learning coverage does not reflect the specification.
- There are inconsistencies in delivery of assessment, notably
    - procedures not followed
    - lack of invigilation staff training
    - insufficient contingency arrangements.

The compliance rate indicates that the majority of centres are implementing assessment to the required standards but that, overall, it is inconsistent. By addressing the weaknesses identified, the standard could be further improved in all centres.

4. Recommendations

The following recommendations are based on the weaknesses noted previously.

- Awarding body representatives should visit centres on a regular basis.
- A variety of teaching methods should be encouraged.
- Clear guidance relating to invigilation procedures should be issued and the implementation monitored. It is recommended that the awarding bodies consider the benefit of agreeing a common approach.

- Guidance in relation to candidates with additional needs must be made clear and where possible consistent across awarding bodies.
- Awarding bodies should monitor more closely the content being delivered to candidates and what is included in assessment.
- Awarding bodies should make the appeals process clearer to centres.

Other issues for consideration

There are two further aspects that the awarding bodies should consider.

The first applies to all awarding bodies. With the exception of one centre, all examinations were taken on the same day that teaching took place. It could be claimed that results in the testing only of short-term memory. Personal license holders should, ideally, be retaining this information for use in their work.

The second issue relates to one particular awarding body. The scrutineers were concerned about the efficacy of immediate re-sits. This is clearly very convenient for candidates and centres. However, if a candidate passes a re-sit, it may show that either the first or the second examination was unreliable as an assessment method. This is because no further teaching/learning or revision has taken place.

PART TWO

5. A comparative evaluation of personal licence holder assessment materials

5.1 Introduction

These National Certificates are new qualifications. Although the awarding bodies concerned are very experienced in developing and running tests of a similar kind, any new qualification faces considerable difficulties. Among these are: ambiguities or lack of detail in the general specification that guides the tests; the need for new question writing teams; the lack of model questions and test papers to emulate; and the absence of clear procedures for setting and maintaining standards. In particular, it takes time to build up the collection of good quality test items needed if awarding bodies are to benefit from the opportunities that effective item banking can bring.

It is against this background that this evaluation of the measurement properties of the assessments of candidate personal licence holders was carried out. The reviewer has considerable experience of test development, item writing and the academic study of test questions, over a period of more than thirty years, and it is, therefore, not surprising that the report contains many comments that are critical of awarding body practice. The intention, however, is to be positive, and the criticisms should be taken as suggestions – just occasionally as demands – for improvements that the awarding bodies can easily implement in order to raise the overall standard of assessment for these qualifications.

5.2 General methodology and materials

This part of the study addresses the comparability of the personal licence holder qualifications in terms of the **quality**, **difficulty** and **cognitive demand** of the multiple-choice tests. In the first sections *quality* is assessed by looking at several indicators of good practice in writing items and constructing tests from them for qualification purposes. Following this is a report on an empirical exercise that estimated how difficult it would for a candidate to answer each question correctly. This provides an estimate of the relative *difficulty* of passing the tests.

Cognitive demand is considered next. That is, the nature of the thought processes that are required of a candidate to answer the questions. This is examined to discover whether any of the

tests demands more sophisticated thinking and mastery of the content than another. The next section reports the results of telephone interviews with representatives of the awarding bodies about the procedures followed to produce the items and tests and to ensure standards are maintained. Finally, the conclusions and recommendations resulting from all of the investigations are summarised.

Three awarding bodies currently set tests for this qualification: the Graded Qualifications Alliance (GQAL), Education Development International plc (EDI), and BIIAB. One set of tests issued by each awarding body was studied, together with relevant documents.

GQAL:	Paper Number 10018	
EDI:	Unit 1 – Licensing Law	ASE 0460 0502
	Unit 2 – Legal and Social Responsibilities	ASE 0461 0506
BIIAB:	Paper Number 4510	

Notes:
None of the test papers carries a precise date (the GQAL test is '© 2005'), but all of the papers appear to have been compiled and administered during 2005.

Except where the discussion concerns the examination format, in this report the two EDI tests are combined and considered as one.

6. Test content and validity

6.1 Specification coverage

The DCMS lays down an outline specification with 67 content sub-topics arranged in 14 topic groups (two of the groups are called *Personal licences*): the full list is attached as Appendix 2 to this report. For this analysis the topics are numbered one through to 14, and a 15th is added for the few items in the GQAL test that do not belong to any of these groups. The topics (with number of sub-topics) are:

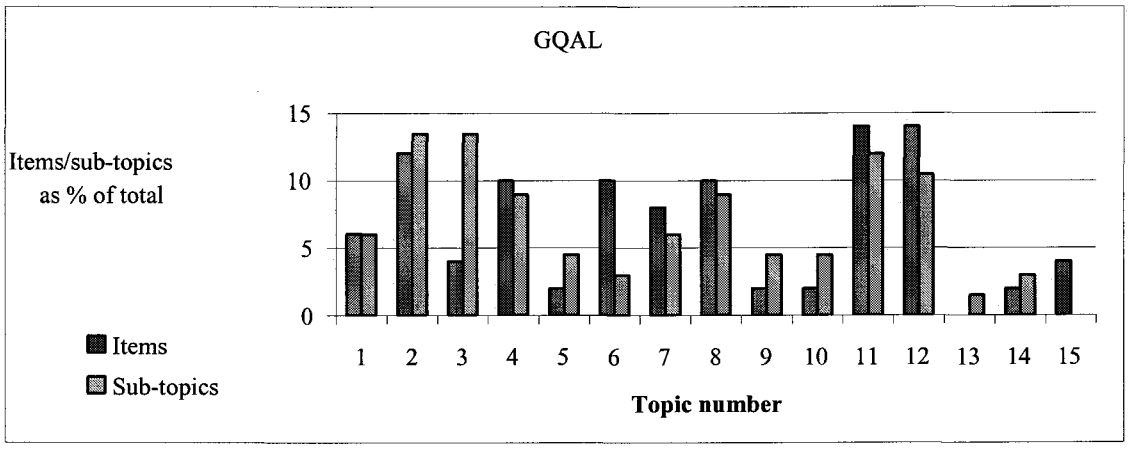
	Topic	*Sub-Topics**
1	Personal licences	4
2	Licensing authorities	8
3	Personal licences	8
4	Alcohol	5
5	Unauthorised licensable activities	3
6	Police powers	2
7	Duties of the personal licence holder	4

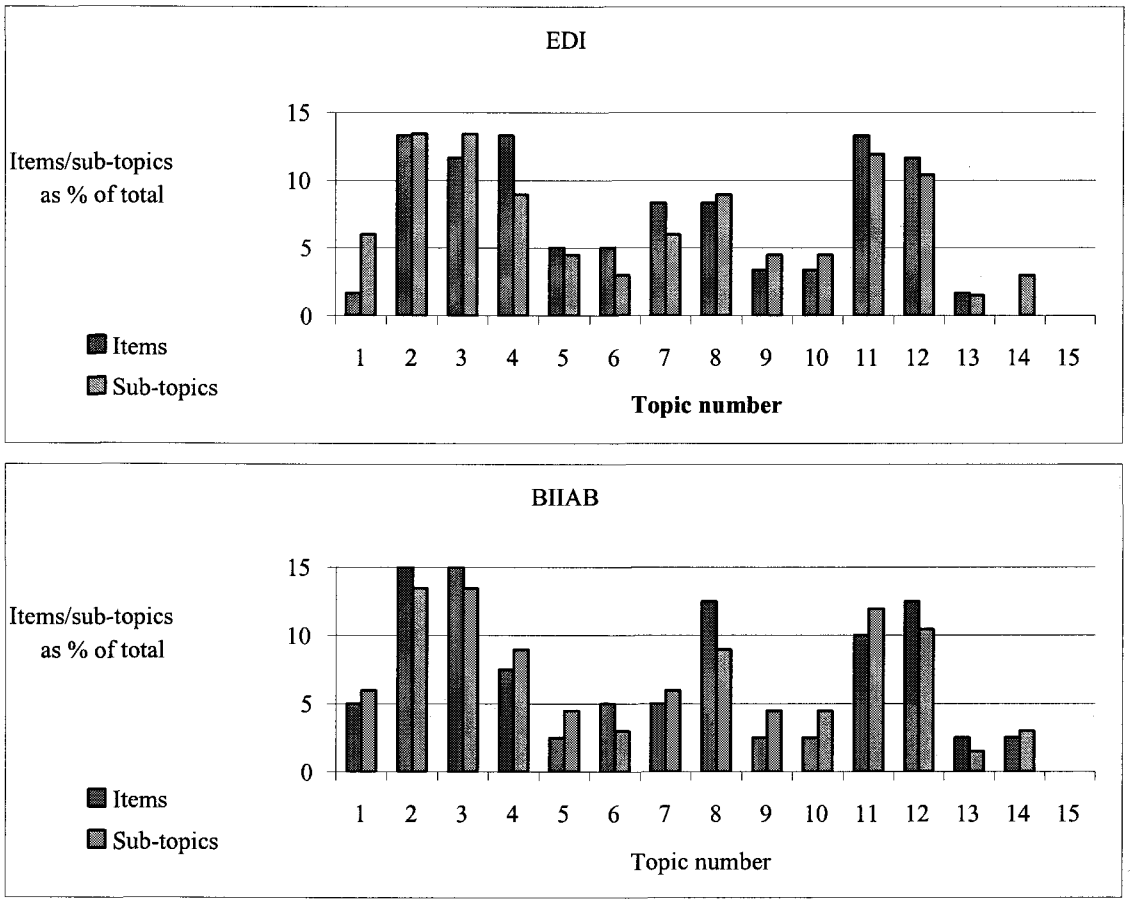
8	Premises licences	6
9	Operating schedules	3
10	Permitted temporary activities	3
11	Disorderly conduct on licensed premises	6
12	Protection of children	7
13	Rights of entry	1
14	Prohibitions	2
15	Other	-

* *Taken from* DCMS website November 2005–March 2006

The three awarding bodies differ somewhat in how they convert this into a test specification, but all of them seem to do this conscientiously, and to keep their own specification under review. They vary in the number of items included – 50 in GQAL's test, 60 in EDI's and 40 in BIIAB's – but this difference is probably not important in terms of coverage; any test can only contain items from a sample of the 67 sub-topics, and some sub-topics will deserve more items than others. The graphs below show how the items are distributed across the topics. In each graph the dark columns show the percentage of test items that relate to each topic, while the light columns show, for comparison, the percentage of sub-topics in that topic in the DCMS specification.

Figure 1: The tests' coverage of content





Interpretation

In all tests the item distribution matches the number of sub-topics fairly closely. There is a general tendency to give a little more emphasis to Topics 6, 11 and 12 – *Police powers*, *Disorderly conduct*, *Protection of children* – and GQAL and EDI ask less about the *Personal licence* itself. These deviations are, however, very small. The question is, rather, whether there should be more variation from the pattern of sub-topics; is it right that every sub-topic should be considered equally important?

It seems that the DCMS list of topics is being used as if it were a specification for test content, but it is not obvious that it was intended to be used in this way. Either DCMS or the awarding bodies should explicitly make judgements about the relative importance of each topic, rather than allowing test content to be decided by default. Furthermore, rigid adherence to any content specification is likely to make text construction more difficult and result in poorer quality tests.

6.2 Questions within topics

There is, of course, more variation within topics. A content specialist would be needed to comment in detail on how the awarding bodies choose to assess each topic. A few issues have arisen, however, during the empirical task in which four judges (all with general experience of the school/further education/higher education system but with no specific knowledge of the personal licence holder qualifications) assessed the relative difficulty of individual questions (see Section 8, *Empirical analysis of perceived difficulty*).

First, several questions asked for factual information at a level of detail that seemed excessive to the judges. For example, candidates were asked to choose the amount of a fine for a particular offence, and to distinguish between (c) *£15,000 fine and/or 6 months imprisonment* and (d) *£20,000 fine and/or 6 months imprisonment*. A representative of one of the awarding bodies reported that the intention here is to impress on candidates the seriousness of the offence. (In fact, as a test strategy they would be well advised always to choose the heaviest penalty in any question like this.) However, both options seem to rate as pretty 'serious'. The Act can be checked easily and quickly, by anyone with internet access, to determine the exact amount of a fine, should this be needed. Is it, then, appropriate to expect candidates to memorise information that is never likely to be needed instantly, and is easily checked when it is needed?

The second point relates, again, to information that may appear unnecessary. For example, some questions on Topic 8 – *Premises licences* – asked much more than it seems reasonable to demand that a *personal* licence applicant should know. There is, of course, no test for applicants for a premises licence, but this qualification is for a *personal*, not a *premises*, licence. It may not be appropriate to ask them if *An application for a review of a premises licence . . . can be made* (a) *Annually* or (b) *At any time*, especially when the question turns on a rather fine legal distinction.

Finally, some questions seemed inappropriate in other ways. For example, all of the tests ask about the meaning of a unit of alcohol, usually quoting the definition in terms of a mass or volume of pure alcohol. This method of measurement has little meaning to most people. Even professional chemists rarely use ethanol that is entirely free of water. Licence holders will deal with various drinks containing varying concentrations of alcohol, and it would be more sensible to ask questions about the amount of alcohol, measured in units, in real alcoholic beverages.

Once a qualification has been in place for two years or so, it would be worth conducting an expert review of the validity of the assessment with regard to issues like these.

All of the tests use exclusively four-option, multiple-choice item format, and all items are the simple selection type (for example, there was no multiple selection or matching and there were no questions offering 'All/None of the above'). In this context this is probably the right strategy.

7. Analyses of question difficulty and quality

7.1 Note on methodology

All of the analyses in this section are subjective, and a different judge would be likely to find different total numbers of cases of each problem in the items. The purpose of the analyses is, however, comparative, in that they should show fairly conclusively which test contained most and which fewest examples of each feature judged.

In order to ensure that a common standard was applied across tests the following procedure was followed:

- Analysis of one feature was completed before the next was begun;
- The first quarter or so of the items in one test were analysed, then the first quarter of another, then the first quarter of the third;
- This was repeated with the second, third and fourth quarters;
- The whole cycle was then repeated for another feature, taking a different test first each time;
- Usually, the early judgements were reviewed at the end of each analysis to ensure that the judge's standard had not changed during the analysis.

7.2 Option plausibility

In a multiple-choice test the difficulty of questions is a function both of the *intrinsic* difficulty of generating/identifying the correct response to the question stem and the *interactive* effect of the other options. The intrinsic difficulty will be addressed more directly in section 8. This sub-section looks at the role of the distractors.

Ideally, all of the options will be plausible answers to the question for candidates who are wholly ignorant of the test content, but only one will be plausible to a candidate who has mastered all of the content in the specification. If an option is not plausible to a candidate it cannot tempt them away from the right answer and thereby help measure their knowledge. For these analyses

reviewers categorise weak options in three ways. (Note: these analyses are three variants of what is described in the literature as Nedelsky's Method.)

The first, *Possible*, imagines a candidate who has some reading difficulty and so may not completely understand the question but can spot give-away clues linking the stem to the correct option. Or there can be obvious language errors that disqualify a distractor: good questions should not have clues like these.

The second, *Plausible*, forecasts the reasoning of a candidate whose reading ability is good, who knows nothing about the relevant content beyond the most basic general knowledge, but who can spot a 'silly' answer. The purpose is to indicate how many options the reviewers may discount on reading comprehension grounds alone, such as when an option does not address the question or gives a clearly non-legal answer to a legal question.

In the third, *Likely*, the author of this review used his many years of experience in writing and scrutinising test questions and in working with question-writers to eliminate unlikely options. These judgements were made *before* he had checked the marking schemes or any relevant content documentation, and in the few questions where he had picked up some specific information he ignored it. The purpose of this variant is to see how much test wiseness and general knowledge could raise someone's score above chance level.

The results of these analyses are converted into test scores as follows: the score on each item is the chance probability of getting it right after options considered 'Impossible' (or whatever) have been removed. Thus, for example, the item score is 0.33 if one option is considered impossible in a 4-option item, or 0.5 if two are. The results are reported in percentages, and the ideal chance percentage is included as a baseline.

Table 23: Option plausibility and test score

Test	Number of items	Ideal %	Possible %	Plausible %	Likely %
GQAL	50	25	26	36	72
EDI	60	25	26	36	70
BIIAB	40	25	25	32	63

Interpretation

In the GQAL and the EDI tests there were a few options judged 'impossible' (2 and 5 respectively). From a total of 200 and 240 options this is not many – they only raise the 'possible' score from 25 per cent to 26 per cent – but there really should not be any in a good test. There were none in the BIIAB test.

A person able to read and understand the questions would be able to raise their score to 32 per cent on the BIIAB test, and to 36 per cent on the others. This means that the other tests gave more irrelevant clues, mostly by including more options that were either implausible or obviously correct. For example:

1 Implausible:

Who has a right of entry into licensed premises? D persons under exclusion orders

2 Obvious:

A Personal Licence renewal must be accompanied by: A a completed application form

The pattern is similar for options judged 'likely': the BIIAB questions contained fewer of the clues that a test-wise candidate could use to get correct answers without the relevant knowledge. Such a person would have passed the GQAL test comfortably, just passed the EDI test and just failed Part 2 of the BIIAB test *without any specific test knowledge*. Of course, they would not have passed BIIAB Part 1 either.

The general conclusion therefore is:
at all levels of test-taking skill GQAL's and EDI's tests had more, inappropriate, plausibility clues.

7.3 Length of options

One of the well-known faults in multiple-choice questions is to make the key option stand out by being a different length from – usually longer than - the distractors. The tests reviewed contained some items like this, and they were counted. The criteria used were:

- the key has more [fewer] words than all the distractors
- **and** the key extends to 2 lines [1 line] with all distractors on 1 line [2 lines]
- **or** the key is at least 50 per cent longer [33 per cent shorter] than the average distractor
- **and** the length differences reflect differences in syntactic complexity:
- **and** there is no good justification for the difference (eg they are all titles of Acts).

An example will illustrate these criteria:

25 Which of the following activities are relevant to a personal licence holders responsibility to deal effectively with disorderly conduct?

A Keeping toilets clean

B Replacing missing bulbs

C Keep the premises generally clean and tidy

D Detection of events building up and acting to prevent the outbreak of violence

The key, D, is the longest option. At 13 words it is three times as long as the average of the others. A simplified phrasal group code[†] was used to indicate syntactic complexity. Here it gave:

A: V.N.A

B: V.A.N

C: V.N.A

D: N.V.V.V.N.N

showing that D was more complex, especially in containing three verbs.

There are other faults in this question – a missing apostrophe in the stem, a change from verb participles in options A and B to a verb in C and a noun in D.

Table 24: The number and percentages of items with length clues

Test	Number	%
GQAL	0	0
EDI	4	7
BIIAB	3	8

These clues will have made the EDI and BIIAB tests a little easier than intended.

7.4 Reading difficulty

Potential reading difficulties were assessed at two levels:

- The first level ('average') is intended to represent an average 17-year-old level 2 candidate. Questions judged problematic at this level were unnecessarily complex or linguistically faulty, in a way that would leave such a candidate uncertain about the intended meaning.

- The second level ('minimal)' is more severe. It is intended to represent a minimally literate person. This level would include many non-native speakers who have reached the Council of Europe's 'threshold' level in English.

[†] V = verb phrase; N = noun phrase; A = adverbial/adjectival phrase. In some other questions additional codes were used, such as C to indicate a complement phrase, or G a negative. The exact form of this analysis was varied to fit the particular question but it was always consistent for the various options within a given question, which is what matters for the analysis in this section.

Table 25: The numbers and percentages of items judged problematic at these levels

Test	Number of items	Reading – average level	Reading – minimal level	Average level (%)	Minimal level (%)
GQAL	50	6	13	12.0	26.0
EDI	60	2	11	3.3	18.3
BIIAB	40	1	7	2.5	17.5

Interpretation

A quarter of GQAL's questions, and between a fifth and a sixth of EDI's and BIIAB's were judged to be problematic for minimally competent readers. This represents a serious literacy problem, especially for candidates who have not been educated primarily in English to GCSE standard. A few questions, especially in the GQAL test, contained language that would be unreasonably difficult for most candidates.

The most frequent cause of these difficulties was awkward, unnatural phrasing. Consider this example:

> What is the most important issue that a badly managed drink's promotion can lead to?

The 'theme' of this sentence is *issue*. Transformation into a question has highlighted *What* as the subject of the sentence and buried *issue* in the middle. This is not in itself difficult, but:

- the theme is complex – *the most important issue*, which requires comparative judgment;
- the theme is qualified by a defining relative clause;
- this clause has a qualified subject – *a badly managed drink's promotion* – with a qualified qualifier – *badly managed*;
- and a complex subject head – *drink's promotion*;
- with a punctuation error – an inappropriate apostrophe;
- and a syntactic ambiguity – is it the *drink* or the *promotion* that is *badly managed*?
- finally, the verb in the relative clause is both modal – *can* (difficult for non-native speakers) – and metaphorical – *lead to*.

In addition, three of the options use non-finite participial verb forms – *not being*, *having*, *drinking* – which are, again, tricky for non-natives (the fourth has no verb at all).
An educated native speaker of English will not find it difficult to construct the intended meaning of this question, but it could have been expressed in a way that would greatly reduce the reading demand:

Bad management of a drinks promotion can cause several problems. What is the most important one?

A mitigating factor for this qualification is that the awarding bodies must test candidates' knowledge of the law, mostly in The Licensing Act (2003), which is not written in language appropriate for a Level 2 qualification. The tests, however, should test the meaning rather than the letter of the law, and more effort should be made to avoid using formal legal language.

7.5 Text highlighting

In all forms of assessment the questions are meant to convey to candidates, as simply as possible, the task they are meant to show they can carry out. The question should not 'get in the way'. Candidates can be helped to understand the question by making its language as natural as possible, but there are other strategies that can also help – the simplest is to use either a bold or an italic font to highlight key words. In these tests:

- GQAL makes this difficult by printing every question in bold, never using italics, and putting just two technical words in single quotation marks. The only consistent emphasis is the capitalisation of NOT (twice).

- EDI prints in a regular font, puts 10 technical terms in single quotation marks (Unit 1) and three technical terms in double quotation marks (Unit 2), does not capitalise any words, and emboldens **not** on the one occasion it is used.

- BIIAB prints in a regular font, puts two technical terms in single quotation marks, and emboldens **minimum (x2)**, **purchase (x1)**, **maximum (x2)**, **consume (x1)**, **must (x4)**, **not (x4)** and **best (x1)**.

The BIIAB strategy of highlighting key words helps focus candidates' minds on the key feature of the question. GQAL should change the print format of their tests.

7.6 Other aspects of question quality

a b c d distribution

Good practice requires test constructors to ensure that each option – a, b, c and d – is used as the key more or less equally often, so that candidates will not use, or attempt to use, inappropriate guessing strategies. In these tests the key was distributed as follows:

Table 26: Distribution of key response across options

Test	a	b	c	d	Total number of items
GQAL	15	12	9	14	50
EDI	17	14	8	21	60
BIIAB	13	10	7	10	40

In every test, option C was correct much less often than the others, and options A and D were most often correct. Part of this effect came from the habit mentioned earlier of setting questions that ask the penalty for some offence – in which the correct answer was always D.

All of the awarding bodies should take more care to spot simple technical problems of this sort before the test goes live.

'Checkable' questions

As described earlier, some questions test knowledge of facts that a licence holder is unlikely ever to need to know instantly, and which are relatively easy to check should the information ever be needed. In these tests the following numbers and percentages of 'checkable' questions were noted:

Table 27: The number and percentages of 'checkable' questions

Test	Number	%
GQAL	5	10.0
EDI	3	6.0
BIIAB	3	7.5

There is no significant difference here between the tests.

Language errors

A number of items contained language errors. These varied from simple typographical slips, such as *before hand* for *beforehand*, to the common misuse or omission of apostrophes, and treating *premises* as a singular noun (in two tests), to grammatical errors that made one or more options ineligible as answers, as in the example:

> To claim damages against police following a closure order a claimant must prove
>
> D where the police only issued a closure order because of lack of police numbers.

The number of errors, and percentages, in each test was counted.

Table 28: Number and percentages of items with language errors

Test	Number	%
GQAL	9	18.0
EDI	7	11.7
BIIAB	3	7.5

Errors in spelling, punctuation and grammar at least convey a bad impression of the awarding body involved. On some occasions they will influence candidates' selection of options, usually with the effect of making the question easier than intended. Item vetting and proofreading should catch most problems of these types. There should be no such errors in the final tests.

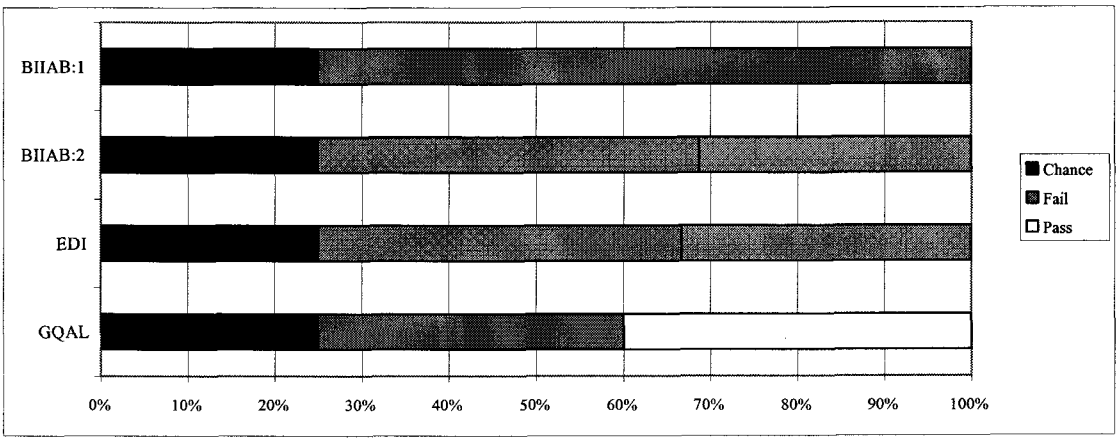
The analysis shows that BIIAB is more successful than EDI, who is in turn more successful than GQAL, at catching language errors. On the other hand, GQAL is to be commended for the clear and natural everyday language that it uses in the questions – when it is not quoting legislation.

8. Empirical analysis of perceived difficulty

8.1 Pass marks of multiple-choice tests, and chance

It should be borne in mind that there is a base-line, the guessing level, which any candidate could be expected to score by chance alone. These tests use only four-option items, with a constant chance of 1 in 4 of being correct by pure luck. The graph below shows the amount of knowledge, in percentage terms, candidates must demonstrate to pass.

Figure 2: Amount of knowledge required to pass



The black strips show the chance effect. The grey strips are the important ones: they show the amount of knowledge, in addition to the chance score, that is demanded from candidates. The GQAL pass mark, of **25 per cent + 35 per cent**, will be less safe – even if the questions are more

difficult – than the EDI pass mark of **25 per cent + 41.67 per cent**, or BIIAB's of **25 per cent + 43.75 per cent + all of Part 1**.

In raw mark terms, beyond the chance score:

GQAL demands **17.5** extra marks

EDI demands **25** extra marks

BIIAB demands **15** extra marks + the **6** extra marks in Part 1

The absolute binomial probabilities of someone passing purely by chance – with absolutely no knowledge at all – are (rounded):

GQAL 1 in **6 m**illion

EDI 1 in **76 b**illion

BIIAB 1 in **272 b**illion,

showing again that candidates are more likely to pass the GQAL test through luck.

More realistically, consider two candidates who definitely know 50 per cent and 40 per cent respectively of the items, and guess the rest randomly: what chance will they have of passing?

Table 29: Probability of passing with low levels of knowledge

% Known	Chance of passing (%)		
	GQAL	EDI	BIIAB (Pt 2)
50	78.6	19.7	19.0
40	19.7	0.9	2.6

Knowing half the items, a GQAL candidate is 80 per cent likely to pass, compared to having just a 20 per cent chance in the others. The one who knows just 40 per cent still has a 20 per cent chance of passing GQAL, but just a 1 per cent or 2 per cent chance of passing the others. (Remember, too, that BIIAB's higher requirement for Part 1 makes it more difficult than this to pass the BIIAB test.) This calculation (before taking into account the effect of question difficulty or demands) suggests the GQAL pass standard should be raised from 60 per cent to about 70 per cent.

8.2 Investigation of question concept difficulty

In most empirical studies of comparability, actual candidate performances are compared. For multiple-choice tests this requires reasonably large samples of candidates (several hundreds) to attempt at least two of the tests being studied, or their own test together with a highly reliable reference test. Neither approach was possible here. Instead, a procedure was adopted that

combines elements of the standard setting procedures commonly used for multiple-choice tests (such as 'Angoff' or 'book-marking' methods) with the paired comparison methodology generally used in comparability studies of GCSE, GCE and GNVQ standards.

Four judges, all with general experience of the school/further education/higher education system but with no specific knowledge of these qualifications, were given seven sets of questions from the three tests. Each set contained four questions randomly selected from each test and the judges were asked to sort the 12 questions into a rank order of relative difficulty, using their judgement of how difficult each question would be if they had just taken a relevant training course. Then, two new sets of 14 questions were assembled, each containing two questions from each of the original sets, and the judges were again asked to rank the questions by difficulty: this allows all of the data to be merged into a single data set. Because the judges report *relative* rather than *absolute* difficulty the method is not affected by any variation in how many questions they think the candidates 'ought' to get right – that is, their data do not judge the standards, but just the relative standards of the tests. Experience with similar methods in investigating general qualifications suggests that the use of non-experts as judges is not inappropriate and may be, in some respects, better than using teachers or examiners.

A general problem with this procedure, however, is that the judges cannot accurately imagine how difficult the questions will be for real candidates. While the comparison methodology 'cancels out' all errors in the standard the judges expect, they may think some kinds of question *relatively* more difficult than the real candidates actually find them. If the tests differ in their use of certain types of question this may lead to inaccuracies in estimating their actual relative standards. This issue of the difficulty of particular kinds of questions will be addressed several times in the report, both from the perspective of empirical difficulty and of validity. ('Empirical difficulty' is defined in terms purely of how many people get the question right, irrespective of what makes it difficult to get it right.)

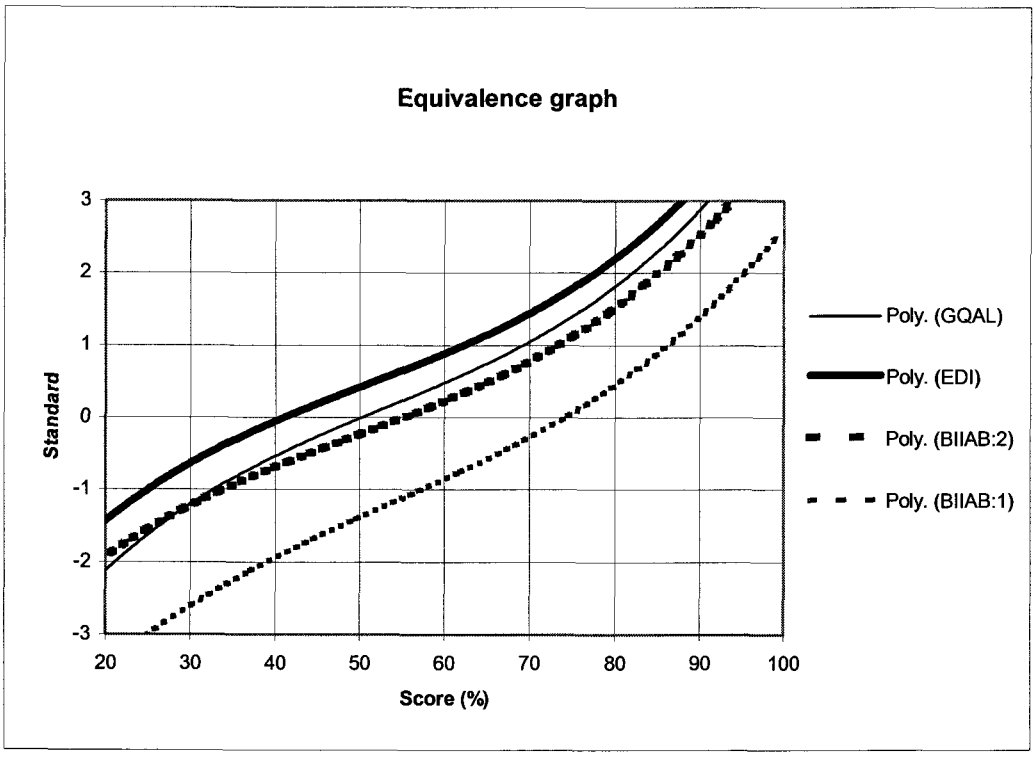
Comparative data of this kind are analysed in line with Thurstone's *Law of Comparative Judgement*, using Rasch measurement theory. This produces a scale of difficulty on which every item used in the exercise is located; selecting just those from each test allows a graph to be drawn relating 'score' to 'standard'. Since the exercise used 28 questions from each test, this graph is re-scaled into percentage score for comparison between the tests. The result is shown in the next graph.

To report the apparent standards of the tests the GQAL and BIIAB pass marks will be compared to EDI's; this should **not** be taken to imply that the EDI test standard is in any way 'correct'.

The table below the graph shows the procedure: the EDI pass mark is first converted (via percentages) to a 'standard' from the graph by reading vertically from 66.67 per cent, and the

corresponding percentages from the other tests are then read from the graph by tracing horizontally to the other lines and then reading down to the percentages. These percentages are converted to give raw score marks equivalent in standard to the EDI pass mark. Note that this procedure is carried out separately for Sections 1 and 2 of the BIIAB test, because different pass criteria are set for them.

Figure 3: Score v Standard for each test



EDI pm = 40%, or 66.67%, which means that the 'EDI standard' = 1.245

Table 30: Differences in difficulty

	Equivalent %	Equivalent score	Actual pm	Difference
GQAL	72.8	36.4	30	– 6.4
BIIAB:2	76.7	24.5	22	– 2.5
BIIAB:1	88.7	7.1	8	+ 0.9

Interpretation

1 Compared to EDI the GQAL test was judged to be considerably easier, by about 6 marks.

2 Compared to EDI the BIIAB test looks easier by between one and two marks but this slightly underestimates the actual difficulty of passing it, because of the 'compulsory question' hurdle: you cannot pass by getting *any* 30 questions right. It is impossible to quantify the effect of this hurdle but reasonable to assume that the BIIAB standard is

overall about the same as the EDI standard. (See Section 9.2 for a further discussion of the effect of hurdles.)

3 It is difficult to gauge the accuracy of this analysis, but experience suggests that a difference of more than about three marks may be considered real.

9. Demands, difficulty and validity

9.1 Cognitive demands

In assessment, the level of cognitive demand refers to the nature of the cognitive processes required of candidates in the process of answering the questions. It does **not** refer to the amount or nature of study required in preparation for the examination, nor to the amount of effort or time the examination takes.

For this exercise a five-level scale was created, with the levels defined as follows:

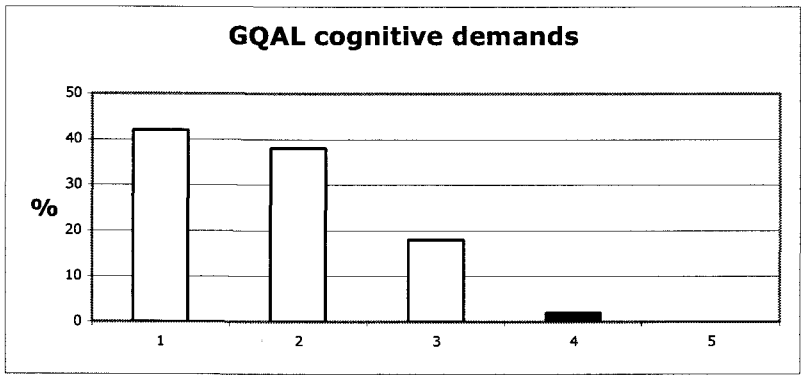
1 simple fact recall, OR simple logic, OR complex recall made easy by options
2 complex recall, including definitions
3 show understanding of a meaning: simple options, OR complex recall made difficult by options
4 show understanding of a meaning: complex options
5 apply reasoning with knowledge, OR show understanding made difficult by options.

The most important difference, for a qualification like this, is between 2 and 3. Levels 1 and 2 test a candidate's ability to recall facts more or less verbatim from a given source, and could be considered to constitute a test only of memory. When testing follows teaching as quickly as it does with these qualifications, these levels principally test short- to medium-term memory. Levels 3 and 4, in contrast, require the candidate to show understanding of what has been learned, by recognising statements equivalent in meaning but expressed differently. Level 5 requires candidates to show that they can exploit the understanding they have achieved in realistic contexts, usually by recognising which of the alternatives offered corresponds in meaning to the principle they were taught.

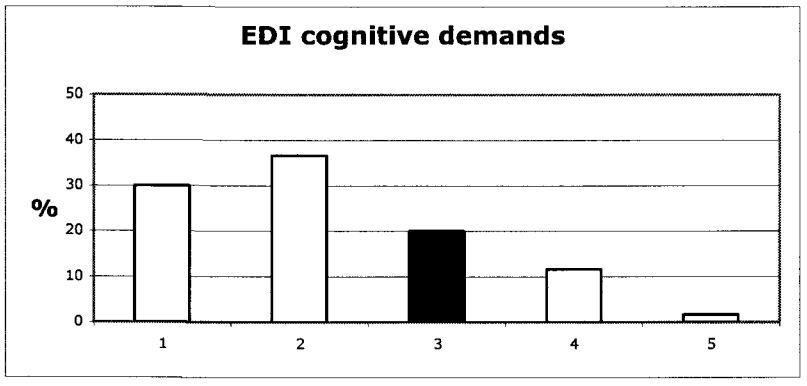
In simple terms, levels 1 and 2 represent a test of memory, while levels 3 and 4 represent a test of understanding, and level 5 a test of application.

The ratings for each test are shown below.

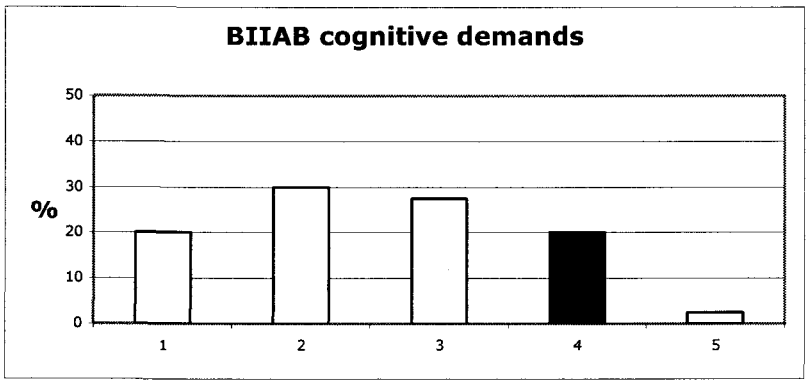
Figure 4: Levels of cognitive demand



Mean demand rating = 1.80



Mean demand rating = 2.13



Mean demand rating = 2.55

The variation in the mean ratings was tested for statistical significance.

Table 31: ANOVA table for demand ratings

	DF	SS	MS	F-Value	P-Value
Test	2	12.590	6.295	6.387	.0022
Residual	147	144.883	.986		

Interpretation

The difference between the tests is highly significant: the probability of it being a chance effect is about 1 in 500.

Thus there are significant differences in the level of *cognitive demand* made by the three tests. Items rated 3 or above (ie understanding and application of understanding) account for half of all the items in the BIIAB test, one third of the items in the EDI test, and one fifth of those in the GQAL test.

The BIIAB test makes the most cognitive demand and the GQAL test the least. Most (80 per cent) of the GQAL items test the recall of knowledge while only half of the BIIAB questions do.

The authorities who oversee this qualification should consider the difference between the three tests carefully. Is this qualification meant to ensure that personal licence holders know the Licensing Act (2003) or that they know what it means? Is it more important to know the definition of "relevant offence" or the potential consequences of committing one? There are, of course, many facts and basic definitions that all licence holders ought to know, but there are also many provisions of the Act whose significance they ought to appreciate and where the ability to reproduce the words of the Act are inadequate proof of understanding. The GQAL items, generally, test knowledge rather than understanding and this imbalance should be redressed. The EDI tests should also shift the balance a little toward testing understanding and application.

One point deserves repeating. When information is taught to candidates who are then tested on it later the same day there can be little guarantee that much of it will be remembered later. In this respect the validity of all the tests, but especially the GQAL test, is questionable.

It is interesting to compare this difference in cognitive demands with the conclusions on difficulty in Section 8. The equivalence line graph there showed that the EDI items were judged the most difficult and the BIIAB ones easiest. Combining the two results implies the following:

GQAL: the items mainly test the recall of knowledge, especially of simple facts. The empirical judges felt that many of these items would nevertheless be difficult to get right, because the answers would be difficult to remember. The pass mark is low, however, making the GQAL test quite easy to pass – unless the candidate finds it difficult to memorise facts and texts.

EDI: the test is intermediate between the GQAL and BIIAB tests with respect to cognitive demands, but the questions tend to be difficult, making the test overall quite difficult to pass.

BIIAB: many of the items tend to assess candidates' ability to understand the content, and to apply that knowledge, rather than just the ability to recall facts. The empirical judges found these comprehension and application items rather easier than simple recall, but the high demand, coupled with the high pass marks required, make the BIIAB test the most difficult to pass.

9.2 Test format demand

Assessment procedures differ across the three qualifications. In this section these procedures are compared.

Time, test length and reading

These issues are inter-related. All other things being equal, candidates would doubtless prefer a short test to a long one, but perhaps they would also rather be given plenty of time to complete the assessment than be pressed for time and, in particular, would expect extra time if the test involves a great deal of reading.

The relevant details for each test are as follows:

Table 32: Analysis of test time and reading demand

	Items	Words	words per item	mins	rate (items per min)	rate (words per min)
GQAL	50	1,912	38.24	60	0.83	31.87
EDI	60	2,535	42.25	80	0.75	31.69
BIIAB	40	1,647	41.18	40	1.00	41.18

The word count here is of the total number of words in the stem plus all the options of each question. Any rubrics not part of the questions are not included.

The BIIAB test is the shortest, but makes the most intense demands, both in terms of the rate of answering questions and of the reading required. The EDI test is the longest, and involves the most reading overall and per item, but also allows most time per item. If reading is not a problem for a candidate then this test will be the most relaxed, but candidates who do have difficulty with reading will find it the most taxing. The GQAL test has slightly fewer words per item than the

others, but lies close to EDI on the measures of rate: it will be the least demanding for candidates who find reading difficult.

Interpretation

The GQAL test slightly favours candidates whose reading standard is not high; the BIIAB test favours candidates who are more comfortable with intense reading and thinking; the EDI test favours those who do not like to be rushed. Unless the typical candidate can be described in more detail, and unless the candidature is unusually homogeneous, it is impossible to say that any of these tests is more or less demanding overall than the others.

Pass marks and the 'compulsory' section

One feature that makes the pass standard difficult to ascertain with these tests is their differences in terms of 'hurdles'. The GQAL test has a single, simple, pass mark (pm) currently set at 30 out of 50. The EDI standard consists of two pass marks of 20 out of 30, with no compensation, which means that it is more difficult to pass two tests at pm = 20 than to pass one test at pm = 40. (In general qualification parlance this was the reason for the use of 'indicators' in awarding multi-unit A levels.) The BIIAB standard is more extreme; as well as using two separate pass marks as EDI does, it sets one of these at 100 per cent.

It is impossible to quantify, without extensive empirical data, the effect of hurdles like these, since the effect depends on matters such as the correlation between scores on the different components and their different reliabilities, but the principles behind setting separate pass marks may be considered.

The two-unit structure of the EDI qualification arises from its policy of offering a series of modules of which two are required for the personal licence holder certificate. The effect, on standards, is to make the qualification just a little more difficult to achieve.

The BIIAB two-part structure is different. It derives from a view that the content to be tested can be separated into two sets – content which every personal licence holder *must* know *all* of and content which *most* personal licence holders should know *most* of: these might be called the *essential* and the *important* content. There are two issues. First, is this a reasonable distinction to make and second, if so, what are the consequences of implementing the distinction in this particular way?

On the first issue there are precedents in various fields of educational assessment – notably in medical domains – for seeing some content as more central and essential than the rest.

On the second issue, there are two causes for concern about the implementation. One is that it is difficult to ensure that the difference between being right and being wrong on these questions turns on an *essential* piece of knowledge. An example from the test analysed here has already been discussed earlier in this report: candidates are required to choose between

(c) £15,000 fine and/or 6 months imprisonment
and
(d) £20,000 fine and/or 6 months imprisonment

as the punishment for selling alcohol outside the hours authorised.

While the main point, that unauthorised sales are a serious offence, is undoubtedly important enough to be considered *essential*, it is not obvious that failing to make the correct choice here is necessary evidence that a candidate does not appreciate the seriousness of the offence.

Secondly, there is a psychological impact in setting a 100 per cent pass mark for a test component. High stakes tests are stressful, and an all-or-nothing demand increases the stress, perhaps unreasonably for anxious candidates. Even experts make mistakes, especially under stress.

The BIIAB should consider, as a compromise, a two-part structure that allows candidates to make a single error in the 'compulsory' section. To set pass marks at 7/8 and 32/40 (or even 7/8 and 33/40) would significantly reduce the stress and make the assessment more palatable, without significantly lowering the actual standard of knowledge and understanding demanded – or unreasonably diluting the message about the extreme importance of some of the legal content.

DCMS might also consider mandating such an approach for all awarding bodies, though there are other issues to be considered before taking this line, and it may prefer to leave this to be considered as a marketing issue by the various awarding bodies.

10. Telephone interview findings

Appropriate people in each of the awarding bodies were interviewed by telephone. These people were used to check various issues that had arisen during the analyses, and to explore the procedures that each awarding body used to write and review questions, to construct the tests, to determine pass marks and to monitor question quality. The interviews also included questions about development plans.

10.1 Question writing

GQAL: A team of four GQAL staff with considerable experience of the trade and in training, and a solicitor specialising in licensing law constitute the panel. They write questions and review each other's questions. Further scrutiny is provided by others with trade experience, and questions are then circulated amongst the panel until they are considered acceptable.

EDI: Specification and specimen papers are drawn up by a 'sector consultant'. Test papers are written and reviewed by an Item Writing Team, the members of which have key skills experience and in-house training. These are reviewed by the item writing teams, proofreaders and subject content experts, with reference to EDI standards documents, and are circulated among the team members until they are considered acceptable.

BIIAB: Items are written by a team of about 10 writers, including solicitors specialising in licensing law and reviewed by a team of two experts and two BIIAB staff against assessment outcomes and other quality criteria. The items are then piloted, and statistics considered, before they are added to the bank.

According to the descriptions of the question writing process provided by representatives of the awarding bodies, EDI and BIIAB seem to have the most satisfactory procedures in place for ensuring that the questions they use are of adequate psychometric quality. GQAL question writers and reviewers have considerable experience in training, but do not seem to have received training in the technical aspects of question writing. Only BIIAB pilots test items before they are used; EDI monitors them very closely during the early life of the test.

10.2 Test production

GQAL: Currently 179 items are used in four, 50-item tests (22 items have been deleted). Another 100 or so items are in preparation. There is, therefore, some overlap between tests. The four tests were created by random selection of items, constrained by a set of construction rules, against a detailed specification with 95 sub-sections (this is being reviewed by merging sub-sections where appropriate). The randomly generated test was checked by an officer and modified by replacement to remove problems.

EDI: The aim is to operate with a larger item bank as soon as possible, and to monitor item quality more actively in order to 'retire' items that function poorly.
The current system is a test construction system, rather than an item banking one; the sector consultant has primary responsibility, like a Principal Examiner in general qualifications, for each test. There are currently five live tests – a total of 300 items – with seven tests in preparation. The aim is to have 12 live tests and to regularly replace outdated ones. Once this is achieved, the system can move on to an item-banking basis.

BIIAB: Items go into a bank after review; there are currently about 540 items in the bank, with about 300 being used in live tests. New items are commissioned regularly, especially when any change in law or regulation demands it. A draft paper is generated by computer and checked for consistency and overlaps.

All of the awarding bodies are moving towards fully operational item banking systems. Although they differ, all of the systems are reasonable at present, and all of the awarding bodies are moving towards a similar system of using computer-generated tests from a substantial item bank. GQAL, in particular, needs to generate additional items quickly, and to run a greater number of tests simultaneously, if it is to be able to utilise the potential of item banking to ensure secure tests and standards.

10.3 Pass marks and item/test statistics

GQAL: The pass mark was agreed at 60 per cent by the five members of the panel, with a commitment to review it after 1,500 results were in – it seems likely that it will be revised upwards[1]. Centres are encouraged to report on any difficulties they see in items; this is the primary route through which item performance is monitored.

A detailed analysis has recently been carried out of 251 results on one of the live papers, looking at such things as item performance, error patterns, any bias with respect to centre, sex, ethnicity or interactions between these factors. In future, analyses like this might be used to screen items for the 'permanent' bank.

EDI: The Item Writing Team approved a pass mark for each test as part of its review, using its experience and reference to earlier 'archive' papers. Completed answer sheets are sorted by score within centres and displayed in a spreadsheet for visual checking. The initial check looks for any evidence of malpractice, and the data are then accumulated for a standard item analysis of question difficulty and distractor performance. Item performance

[1] GQAL has subsequently raised its pass mark from 60% to 70% as of 1st July 2006.

statistics are reported back to the Item Writing Team. The pass mark will be moved if the evidence from early results suggests it should be. This process continues until 200 candidates have taken the test. A final review then approves the test for routine use, and the monitoring becomes less intense.

BIIAB: The pass mark for Part 1 is set at 8 out of 8, based on a recommendation to the Secretary of State for Culture, Media and Sport by a stakeholder advisory group, of which the BIIAB was part, that the test specification should require a 'compulsory section' of this kind. In the event, the Secretary of State did not make this a mandatory requirement, but BIIAB chose to retain a compulsory section.

The pass mark for Part 2 is constant across tests, and was agreed by the moderating team following discussions with a wide group of interested parties, including deliverers of training, content experts, test users and magistrates. Feedback is invited from centres. Simple item analysis is run approximately monthly; a report for each item shows how responses are distributed across the key and distractors, and is scanned for evidence of problems.

Given the high pass marks that are demanded, the principal indicator of an item not functioning well is a low facility value – <90 per cent for Part 1 and <50 per cent in Part 2.

No awarding body has a very secure way of ensuring that a constant standard is being set across all of their tests, though all of them apply various checks to try to ensure this. The weaknesses in these procedures arise mainly from the newness of the qualification. The awarding bodies are at different points in the process of moving towards a fully functional item banking system and recognise that they will then need more robust, routine, item quality control procedures.

10.4 Plans for future development

GQAL: Current concerns include the language demand, especially as it relates to non-native speakers. GQAL is developing translated versions of the tests, initially in Welsh and Singhalese. Given the concern noted earlier that the GQAL test was over-concerned with the memorisation of facts, this move to ensure that candidates understand what is being asked is commendable.

Additional modules are being produced as options to accompany the qualification, and a diploma framework is under development. The team needs to generate more items as quickly as possible to enable more functions of an item banking system to be developed.

EDI: Analyses of current data will be run to review the pass marks in the different tests. Once the current test-based system is stable, with 12 simultaneous live tests, a move to a bank-based system is planned. Items can then be replaced in individual tests to adjust the test difficulty to a common standard.

EDI already administers Key Skills tests online, and has piloted online registration, delivery and marking of international tests. It plans to offer the personal licence holder qualification online too.

BIIAB: The current system is a hybrid perhaps best described as a test bank in the process of evolving into an item bank. Plans centre around its development into a full item bank system supporting e-assessment and an internal pilot has been carried out. The awarding body is alert to the potential of a bank for maintaining flexibility as legislation and practice force changes in the tests.

10.5 General comment

There is room for improvement in the procedures of all of the awarding bodies. As the qualification settles down, the maintenance of a question bank of items of adequate quality and a system for generating tests flexibly from it, will become more necessary. Item analysis based on live data will be central in improving the quality of the items used in the tests. A bank supporting many simultaneous test versions, or an on-screen system for delivering different tests to different candidates, is probably the best way to guarantee the security of test papers or on-screen tests.

11. Overall strengths and weaknesses

11.1 Strengths

The following may be considered as strengths in some of the tests analysed:

Content validity

All three of the awarding bodies ensure that their tests closely match a content specification.

Test design

All of the tests used a simple four-option multiple-choice format, which is quite appropriate for this qualification. The strategy of a 'compulsory section' in the BIIAB test is ill-advised from a technical

point of view, but may be defended on other grounds and shows a commendable concern for the wash-back effect of the test on licensing training and practice.

Item banking

All of the awarding bodies are moving towards fully operational item banking systems. Although they differ, all of the systems at present are reasonably good and, as they develop, will be better able to ensure that standards are maintained.

11.2 Weaknesses

Some weaknesses were also noted in the construction of the tests. Attention to these would improve the quality of assessment.

Dubious item validity

In some cases questions were asked that may be inappropriate in a qualification test for personal licence holders.

Language errors

There were many errors in the spelling, punctuation or syntax of items. The awarding bodies differed greatly in this.

Item quality assurance

Training in writing skills is essential, so that the many kinds of item fault noted in this report can be avoided. Adequate procedures for catching faults by review or pre-test were missing in some cases.

11.3 Strengths and weaknesses

In some respects the tests showed both good and bad features together.

Test language

All of the awarding bodies try to control the language used to ask the questions, and show that they are aware of the need to make questions comprehensible to the test takers. Sometimes, however, this means using the formal language of legislation, which is difficult for many level 2 candidates while, in some other cases, unintended syntactic complexity or unnatural phrasing cause invalid difficulty.

Standards

One of the qualifications appears to be significantly easier to pass than the others. This can be remedied easily, and the awarding body in question was already considering a correction when this review was conducted. All of the awarding bodies appear to be actively monitoring results to check test difficulty.

Cognitive demand

The tests vary significantly in their balance between questions that test recall of facts and those that test understanding. It seems clear that this qualification should require candidates to show that they understand the relevant legislation and regulation, and it is suggested that at least half of the questions should address this.

Item statistics

All of the awarding bodies are able to monitor item performance statistically, and do so to some extent. With a more developed item bank in place some of them will need to develop more efficient and automated ways of doing this.

12. Conclusions

- Although they differ slightly in how they do so, all of the awarding bodies ensure that their tests adequately cover the test specification.

- There are some concerns about the nature of several types of questions used. Some seem to place excessive demand on memory – of particular concern when testing immediately follows teaching so that the memory used is short-term.

- Other questions address information that is unimportant or is unreasonable to expect every candidate for a Personal Licence to know. Once a qualification has been in place for two years, it would be worth conducting an expert review of the validity of the assessment to ensure that the testing is meaningful for typical workplace contexts and statutory requirements.

- All of the tests used the four options – a, b, c, d – unevenly as the right answer; they all favoured a and d and avoided c.

- A few options in the EDI and GQAL multiple-choice items were judged 'not possible' right answers. This should not happen. In all of the tests, an intelligent but ignorant candidate

could expect to score around one-third marks – a little less in the BIIAB test and a little more in the other two.

- The length of options gave invalid clues to the right answer in about 7 per cent of the EDI and BIIAB items.

- Reading difficulty was judged to be a serious invalid source of difficulty in a significant number of questions, especially if candidates are not native speakers of English. This was more prevalent in the GQAL test than in the others.

- Question writers need to make more effort to test the meaning of the law rather than knowledge of its actual wording, since legal language is inappropriately difficult for a level 2 qualification.

- Text highlighting was not used by GQAL, and not often by EDI, to help candidates understand the meaning of items. BIIAB used highlighting well.

- There were many errors of spelling, punctuation or grammar, especially in the GQAL test; this spoiled GQAL's commendable attempts to use everyday language to express questions.

- Luck can play too great a part in helping candidates to pass the GQAL test. Empirical investigation suggests that the GQAL pass mark is too low, by about 6 marks. Raising it would reduce the role of luck.

- The BIIAB test makes more intense demands in a relatively short test. The EDI test is the most comprehensive, with a greater number of questions, a greater amount of reading and allows more time than the others. The GQAL test is the most appropriately designed test for candidates who have difficulty with reading English (despite having more items that are difficult to read because of faults).

13. Recommendations

- During the next 12 months a review of questions should be carried out, by or including an independent content expert, to identify items that address inappropriate content or appropriate content in an inappropriate way, and to suggest better item types to replace

them. Assessment validity depends on testing that is meaningful for typical workplace contexts.

- A review, perhaps in liaison with the DCMS might also consider whether all of the topics/sub-topics in the DCMS specification deserve equal weighting. The review should consider whether it is acceptable for candidates to be tested on knowledge on the same day that they acquire it.

- Several of the faults commonly described in textbooks and training materials for multiple-choice item writers were common in these tests. Procedures for catching these should be tightened.

- All of the awarding bodies should be encouraged to move quickly towards their declared aim of implementing fully featured item banking systems. This will improve the quality of the items that make up the tests and increase the confidence that certificate users may have in the standard of the qualification.

- Random selection of items within test 'modules' will provide the best insurance that the standards of parallel forms for these tests do indeed set the same pass standard.

- Systems for monitoring the quality of items in order to identify and delete faulty items (both at item review stage before they are used in tests and through analysis of real test data), need to be improved if the benefits of item banks are to be realised

- The BIIAB test asks more questions that require evidence that candidates understand and can apply their knowledge than the other tests. The GQAL test is the least demanding in this respect, and should increase its use of items that test understanding. EDI should also adjust the balance of its items a little in the same direction.

- The 'compulsory' section in the BIIAB test considerably increases the difficulty of passing the test. It is not obvious that the items in this section are always the ones that we could reasonably insist every candidate must know. It is recommended that the awarding body instead sets a high but not perfect pass mark for this section. The DCMS and the regulatory authorities should consider whether a single policy on this issue should be mandated for all of the qualifications.

Appendix 1: Glossary of awarding bodies

BIIAB – BIIAB

EDI – Education Development International plc (GOAL)

GQAL – Graded Qualifications Alliance

Appendix 2: The DCMS test specification*

Topics and sub-topics

1 Personal licences

What they are

What they entitle the holder to do

Period of validity

Who grants them

2 Licensing authorities

What they are

How they work

Licensing objectives – what they are

Functions of licensing authorities

Importance of partnerships

Role of Crime Reduction Partnerships

Licensing policies

Hearings

Appeals

3 Personal licences

Procedures for application

Criteria for new personal licences and renewals

Determination of application

Persons disqualified from

Penalty for selling without

Convictions during application and after grant or renewal

Relevant offences

Forfeiture or suspension of licence on conviction

Penalty for breach

4 Alcohol

Definition of supply of alcohol

Premises to which the definition applies

Wholesale and retail sales

Nature of
Strength of intoxicating drinks
Alcohol in the body etc

5 Unauthorised licensable activities

Unauthorised sales
Defence of due diligence
Penalties for breach

6 Police powers

Suspension and closures
Antisocial Behaviour Bill – EHO powers of closure (to be inserted when legislative programme complete)

7 Duties of the Personal Licence Holder

Notification of convictions
Changes in name and/or address
Production of licence to authorised personnel
Penalties for breach

8 Premises licences

What they are
Licensable activities and what they are
Definition of regulated entertainment
Role of designated premises supervisor
Need for risk assessment as designated premises supervisor
Awareness and prevention of crime, disorder and anti-social behaviour in and around licensed premises

9 Operating schedules

What they are
What they should include
Children in licensed premises

10 Permitted temporary activities

Definition
Frequency

Police objections

11 Disorderly conduct on licensed premises

Rights and duties of authorised person

Illegal drugs

Relevant offence of drink driving

Prevention of nuisance

Pubwatch/Retail Watch schemes

Portman Group - responsible drinks promotions

Consequences of irresponsible drinks promotions

Penalties for breach

12 Protection of children

Sale of alcohol to and by young persons

Proof of age cards and schemes

Consumption of alcohol by young persons

Test purchasing

Penalties for breach

Defences

Importance of the awareness of other relevant legislation related to the protection of children

13 Rights of entry

Rights of entry

14 Prohibitions

Moving vehicles

Service areas etc

* This specification is dated January 2006