Nunes, C.A.A., Nunes, M.M.R. & Davis, C. (2003). Assessing the inaccessible: metacognition and attitudes. *Assessment in Education: Principles, Policy and Practice*, **10**, 3, 375–388.

Peat, M. & Franklin, S. (2002). Supporting student learning: the use of computer-based formative assessment modules. *British Journal of Educational Technology*, **33**, 5, 515–523.

Ridgway, J. & McCusker, S. (2003). Using computers to assess new educational goals. *Assessment in Education*, **10**, 3, 309–328.

Russell, M., Goldberg, A. & O'Connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy & Practice*, **10**, 3, 279–293.

Thelwall, M. (2000). Computer-based assessment: A versatile educational tool. *Computers and Education*, **34**, 1, 37–49.

Threlfall, J., Nelson, N. & Walker, A. (2007). *Report to QCA on an investigation of the construct relevance of sources of difficulty in the Key Stage 3 ICT tests*. Retrieved February 26, 2009, from http://www.naa.org.uk/libraryAssets/media/Leeds_University_research_report.pdf.

Topping, K.J., Samuels, J. & Paul, T. (2007). Computerized assessment of independent reading: Effects of implementation quality on achievement gain. *School Effectiveness and School Improvement*, **18**, 2, 191–208.

Welch, A.R. (1998). The cult of efficiency in education: comparative reflections on the reality and the rhetoric. *Comparative Education*, **34**, 2, 157–175.

Wirth, J. & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy and Practice*, **10**, 3, 329–345.



# Is CRAS a suitable tool for comparing specification demands from vocational qualifications?

**Jackie Greatorex and Nicky Rushton** Research Division

## Introduction

Historically, unitary awarding bodies and the national regulator[1] monitored standards of qualifications between awarding bodies, over time and between cognate qualifications at the same level, and this work continues. A key reason for conducting such work is to avoid inequalities and inequities which would be created by the existence of easier routes to access further study or jobs.

Ideally standards are compared in terms of candidates' performance and in terms of the demands of the qualifications. When comparing new qualifications there is sometimes a lack of performance evidence[2] or assessment tasks[3] to form a robust sample from which generalisable research results can be drawn. In such cases comparability studies could focus on specifications[4] and the associated demands. However, studies restricted to one aspect of comparability (whether it be performance or demands) are limited.

One approach to comparing demands of qualifications is for experts to rate them on a scale of cognitive demands known as CRAS. CRAS was developed using academic qualifications. An issue deriving from its provenance may be that CRAS is not suitable for use with vocational qualifications which are different in nature and purpose to academic qualifications. Generally there are far more comparability studies about academic qualifications than VQ/VRQs[5]. In the present study we investigate whether CRAS is suitable for use in comparability studies which include VQs/VRQs.

## Demands and difficulty

There is sometimes a lack of clarity about definitions of demands and difficulty.

In this article:

*Task demands* refer to the actions (usually cognitive) a task is intended to require of typical members of the target group of learners. For example, candidates might be required to recall familiar information. Task demands generally relate to individual summative assessment tasks such as examination items. But task demands could also be related to an individual classroom activity or similar.

*Specification demands* refer to the actions the specification is intended to require of typical members of the target group of learners in four areas: cognitive, affective, psychomotor and interpersonal. These specification demands might be explicit in the specification or they might be an underpinning ethos. For example, candidates might be required to recall information about a topic, empathise with another person's understanding of the topic, evaluate the other person's understanding to know what extra information they need and explain the relevant information to the

1. Currently the national regulator of the awarding bodies is Ofqual.
2. *Performance evidence* refers to students' work in the form of essays, artefacts, paintings, multiple choice responses and so on.
3. *Assessment tasks* refers to examination questions, assignments, briefs for work-based projects and so on.
4. The specification is: *The complete description – including optional and mandatory aspects – of the content, assessment arrangements and performance requirements for a qualification. A subject specification forms the basis of a course leading to an award or certificate. Formerly known as a 'syllabus'.* QCDA (undated)
5. VQ refers to vocational qualifications and VRQ to vocationally related qualifications. These are very broad categories. Many vocational qualifications in England are NVQs (National Vocational Qualifications which: *are designed to recognise a candidate's competence in the workplace. They provide a statement to employers of skill, competence and knowledge in a particular sector.* (OCR, 2009a).
Vocationally related qualifications generally focus on an occupation or occupational sector: *Vocationally-Related Certificates enhance knowledge and build upon candidates' skills in preparation for a job.* (OCR, 2009b).

other person in an accessible manner. These examples of specification demands are cognitive and interpersonal. Specification demands relate to the specification; they are not about individual summative assessment tasks such as examination items.

*Demanding* refers to the extent to which a task or specification is intended to be challenging for typical members of the target group of learners.

*Difficulty* refers to "an empirical measure of how successful a group of students were on a question." (Pollitt *et al*., 2007, p.169). Relative difficulty can be measured as facility values; that is, the proportion of candidates giving the correct response to an item (Kline, 1986).

The notion of *intention* is crucial in clearly defining the concepts of task demands and difficulty. Task demands are about what typical members of the target group of learners are expected or intended to do to carry out a task. Difficulty is focused on the students' actual performance. Bloom (1956) emphasises this difference between what is intended and what actually happens in his work to develop a taxonomy of educational objectives.

The definitions of demands and difficulty used in this article are given above. However, there are various definitions of demand(s) which are used by other researchers for different contexts and purposes, for examples see Pollitt *et al*. (2007) or Barry (1997).

Awarding bodies and the national regulator have used various methods to compare the demands of academic qualifications. One approach has been to develop a questionnaire about task and specification demands, which senior examiners use to rate the task and specification demands of the various qualifications, for example, see Edwards and Adams (2003).

### CRAS

Another approach to comparing task demands is to rate examination items on a scale of cognitive demands known as CRAS. The five aspects of the CRAS frame of reference given below are taken from Pollitt *et al*. (2007).

- "*Complexity:* The number of components or operations or ideas and the links between them." For example, using a single idea is less demanding than synthesising several ideas.
- "*Resources:* The use of data and information." For example, using all and only the information provided is less demanding than selecting the appropriate data.
- "*Abstractness:* The extent to which the student deals with ideas rather than concrete objects or phenomena." For example, work which deals with concrete objects is less demanding than mostly abstract work.
- "*Task strategy:* The extent to which the student devises (or selects) and maintains a strategy for tackling the question." For example, when a strategy is provided this is less demanding than when a strategy needs to be devised by the student.
- "*Response strategy:* The extent to which students have to organise their response." For example, giving the student a small number of possible responses to choose between is less demanding than them having to organise their own response.

The text in quotation marks is from Pollitt *et al*. (2007, p.186).

However, various concerns have been raised about the use of CRAS:

1. It has recently been used by QCA[6] to rate whole examinations rather than individual tasks; it was designed for the latter not the former.
2. In the context of comparing VQ/VRQs CRAS may not be suitable as it was developed using academic qualifications (Hughes *et al*., 1998), which can be different in nature and purpose.
3. Whilst it can be used to compare task demands from academic qualifications it may not be applicable to VQ/VRQs specification demands.

In the present investigation the second and third concerns are addressed.

Crisp and Novaković (2009) used CRAS to compare the task demands from different centres for college-assessed units in a VRQ. They found that complexity, resources, task strategy and response strategy could be used to compare the task demands of various vocational assessments in one domain. However, abstractness was of less relevance.

In the present study we investigated whether CRAS was suitable for use in comparability studies about the assessment tasks and the specification of VQs/VRQs. To do this the CRAS frame of reference was compared with the frames of reference used in previous studies that compared the task and/or specification demands of VQ/VRQs.

## Data

Data for the present study were taken from a series of comparability studies by awarding bodies or the national regulator about VQ/VRQs which are in the public domain (SCAA, 1995; Coles and Matthews, 1995, 1998; Arlett, 2002, 2003; Guthrie, 2003; QCA 2006, a and b). The data were the frames of reference used to compare qualifications in various studies about VQs/VRQs. The studies are outlined below.

Arlett (2002, 2003) and Guthrie (2003) used a modified version of Kelly's Repertory Grid to elicit the similarities and differences between VCE[7] qualifications from different awarding bodies in terms of summative assessment and specification requirements. The similarities and differences were used to develop items for a questionnaire on which senior examiners rated the various specifications, assessments, mark schemes or equivalent, and teacher support materials. The ratings were used to compare the qualifications. This approach was used in two vocational subjects.

SCAA (1995) asked subject experts to judge specifications, guidance to centres, examination papers and internal assessment[8] material/ instructions and guidance against a series of factors drawn from the GNVQ[9] grading criteria and an UCLES[10] specification. The factors were:

> *Content: depth, breadth. Skills: factual recall, understanding and explanation, planning, investigation, analysis and evaluation, transferability (including the extent to which the student is encouraged to be adaptable and versatile) and application of skills*.
> (SCAA, 1995, p.4).

Breadth and depth refer to the breadth and depth of the qualification content which was studied and tested. The experts were also asked to judge whether the time requirements of the specification were likely to be met.

6. The Qualifications and Curriculum Authority (QCA) was once the regulator of the awarding bodies. It was a predecessor of the Qualifications and Curriculum Development Agency (QCDA) and Ofqual.
7. VCE or Vocational Certificate of Education is also sometimes referred to as the AVCE Advanced Vocational Certificate of Education. It was intended to replace the advanced General National Vocational Qualification (see below). In September 2005 VCEs were renamed GCE A-Levels (General Certificates of Education) in applied subjects. The specifications aim to give a broad introduction to vocational domains and to facilitate learning, teaching and assessment in work-related contexts. This information is from the Learning and Skills Council (2009).

QCA (2006a) used the following for subject experts to rate the level of cognitive demands of various multiple choice tests:

1. *Simple fact recall OR simple logic OR complex recall made easy by options*
2. *Complex recall including definitions*
3. *Showing 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*. (QCA, 2006a, p.43).

A similar method and the same definitions of each level of cognitive demands were used in QCA (2006b) a comparability study of assessment practice for Door Supervision qualifications.

Additionally, in QCA (2006a) subject experts rated the plausibility of options in multiple choice tests. The reading difficulty of tests was identified and the accessibility of the questions was quantified by noting instances when important text was highlighted, perhaps by making it bold or italic. These issues, whilst they are not labelled "cognitive demand" by QCA (2006a), are similar to some of the items in Arlett (2002, 2003) and Guthrie (2003).

Coles and Matthews (1995, 1998) undertook a complicated methodology to qualitatively compare qualification learning outcomes, aims and content with a frame of reference, rather than compare the qualifications with one another. To create such a measure they adapted Bloom's taxonomy by adding a skills component based on the work of Gagne (1985) and Mitchel and Bartram (1994). Coles and Matthews (1995, 1998) argue that they needed the latter works to ensure that Bloom's taxonomy was not biased towards academic qualifications. Their frame of reference was based around recall, practical capability, interpretation, application, analysis and synthesis. They defined each term for the purposes of their study, then used this new frame of reference to classify the qualification and assessment requirements and to describe the specifications. Once the specification, learning outcomes and aims were classified in terms of the frame of reference the qualifications could be compared in detail.

In summary, the following were used as data in our study:

- The questionnaire items from Arlett (2002, 2003) and Guthrie (2003).
- SCAA's criteria, as well as the issue of time.
- QCA's levels of cognitive demands, plausibility of multiple choice options, reading difficulty and accessibility of text.
- Coles' and Matthews' (1995, 1998) frame of reference.

## Procedure

The authors classified the data into three groups:

1. Referring to one of the five aspects of CRAS.
2. 'Other' (referring to task and/or specification demands not covered by CRAS).
3. 'Not' (referring to something which was not task and/or specification demands).

Initially one researcher classified the data. The judgements were checked by a second researcher and discrepancies were discussed and resolved. It was acknowledged that elements such as the reading difficulty of a test might be classified as more than one aspect of CRAS so multiple classifications were allowed. Examples of some judgements are given in Table 1.

To make and quantify the judgements, the data were divided into units. For some studies like Arlett (2002) each questionnaire item could be used as a unit. Each row in Table 1 represents a unit.

## Limitations

Inevitably there is some subjectivity involved in the unitisation and the judgements, and other researchers might have come to somewhat different decisions. Nonetheless, the present study is a credible way of investigating the utility of the CRAS framework for comparability studies about VQ/VRQs.

## Findings

The results of the study are presented in Table 1 and Table 2. The majority of the data corresponded to an aspect of the CRAS frame of reference.

However, there were some data which did not map to CRAS, but which were classified as a task and/or specification demand(s). For instance, "More general capabilities such as the ability to work in a team" (Coles and Matthews, 1995, p.11), was predominantly affective and interpersonal, whereas CRAS is primarily concerned with the cognitive. Whilst these classifications are assigned to the minority rather than the majority of the data, they are arguably significant in VQ/VRQs. Therefore, using CRAS for comparability studies for VQ/VRQs is likely to mean that some task and/or specification demands, which are significant in VQ/VRQs, are not included in the research.

One of the most striking results is that we did not classify up to 39% of the data from Arlett (2002) as task and/or specification demands. Table 3 provides some data that were classified as not being a task and/or specification demand(s), along with the reason for that decision. Our findings confirm those of Pollitt *et al*. (2007) who found that comparability studies often aim to investigate task and/or specification demands when they are actually investigating something quite different. Indeed it suggests that there is a need to disseminate the technical term and definition of task and/or specification demands to assessment professionals, researchers, assessment setters, specification writers and users of assessments. Otherwise communication can become unclear.

8. Internal assessment is: *A form of assessment where assessment tasks are set and learners' work assessed by the centre, subject to external moderation or verification where appropriate*. (Ofqual, 2008).

   Moderation is: *The process through which internal assessment is monitored to ensure that it meets required standards, and through which adjustments to results are made where required to compensate for any differences in standards that are encountered*. (Ofqual, 2008).

9. GNVQs or General National Vocation Qualifications aimed to provide study for those intending to stay in full time education but who were not deemed able enough for an A-level programme. The specifications included academic education as well as some vocational learning experiences. The assessments were primarily competence based, evidence gathering and portfolio based rather than external examinations. This information is from Savory *et al*. (2003).

10. The University of Cambridge Local Examinations Syndicate (UCLES) now has the brand name Cambridge Assessment, which was not in use when SCAA (1995) was written.

**Table 1: Examples of data from comparability studies and judgements about how they do or do not map to aspects of CRAS**

| *Data and data source* | *Complexity* | *Resources* | *Abstractness* | *Task strategy* | *Response strategy* | *'Other' task and/or specification demand(s) not in CRAS* | *Data 'not' considered to be a task and/or specification demand(s)* |
|---|---|---|---|---|---|---|---|
| "Evaluation: making judgements based on criteria which have been developed for the purpose. Such as the evaluation of the efficiency of a multi step production process" (Coles and Matthews, 1995:12) | ✔ | ✔ | ✔ | ✔ | ✔ | | |
| "How much opportunity is provided for candidates to apply knowledge in their answers to the question paper? A little to a lot" (Arlett, 2002: 12) | ✔ | ✔ | ✔ | ✔ | ✔ | | |
| "How specific is the breakdown of marks in the mark schemes? Less specific to more prescribed" (Arlett, 2002: 14) | | | | | | | ✔ |

Note: A tick indicates that the data correspond with an aspect of CRAS.

**Table 2: Frequency of data from comparability studies about VQ/VRQs that do or do not map to CRAS**

| *Study* | *Total* | *Complexity* | *Resources* | *Abstractness* | *Task strategy* | *Response strategy* | *'Other' task and/or specification demand(s) not in CRAS* | *Data 'not' considered to be a task and/or specification demand(s)* |
|---|---|---|---|---|---|---|---|---|
| Coles and Matthews (1995) | 13 | 11 85% | 11 85% | 11 85% | 11 85% | 11 85% | 2 15% | 0 0% |
| SCAA (1995) | 11 | 9 82% | 9 82% | 9 82% | 7 64% | 7 64% | 1 9% | 1 9% |
| Arlett (2002) | 23 | 12 52% | 8 35% | 7 30% | 8 35% | 8 35% | 0 0% | 9 39% |
| Arlett (2003) | 35 | 18 51% | 17 49% | 11 31% | 6 17% | 9 26% | 3 9% | 11 31% |
| Guthrie (2003) | 26 | 12 46% | 12 46% | 12 46% | 8 31% | 10 38% | 2 8% | 8 31% |
| QCA (2006a) | 8 | 8 100% | 7 88% | 6 75% | 6 75% | 6 75% | 0 0% | 0 0% |

Note: The first column lists the studies which were included in our investigation. The column labelled 'total' gives the total number of units from each study. The remaining columns refer to the classifications – namely the various aspects of CRAS as well as the categories 'other' and 'not'. Each of these remaining columns has two sub-columns, the left hand sub-column indicates the number of units receiving each classification and the right hand sub-column indicates the number of classified units as a percentage of the total number of units in each study. For each unit more than one classification was allowed, and this is why the percentages in each row do not total 100%.

**Table 3: Examples of data and the reason why it was not classified as a task and/or specification demand(s)**

| *Data uni* | *The reason the data was not classified as a task and/or specification demand(s)* |
|---|---|
| "Is the number of marks allocated to each question appropriate?" (Arlett, 2002:13). | Essentially this is an issue of whether the mark scheme was well written and mark allocation was appropriate. The actions a task is intended to require of typical members of the target group of learners are not directly affected by the number of marks allocated to the question. |
| "Does the mark scheme allow for much compensation/ interpretation?" (Arlett, 2002: 14). | This is about style of mark scheme and whether they allow compensation or whether they are criterion referenced. The actions a task is intended to require of typical members of the target group of learners are not directly affected by whether the mark scheme allows compensation or whether it is criterion referenced. |
| "How helpful are the mark schemes to: Examiners, in ensuring consistency in marking?" (Guthrie, 2003: 12). | This is about the utility of the mark scheme for examiners. The actions a task is intended to require of typical members of the target group of learners are not directly affected by whether the mark scheme is helpful in ensuring consistency of marking or not. |
| Whether: "the stated objectives of each scheme were met by the materials considered" (SCAA, 1995: 4). | This is about validity. There are various elements to validity and in this case the issue is the correspondence between what is supposed to be and what actually was measured. The actions the specification is intended to require of typical members of the target group of learners are not directly affected by the correspondence between what is supposed to be and what actually was measured. |

**References**

Arlett, S. (2002). *A comparability study in VCE Health and Social Care units 1, 2 and 5. A review of the examination requirements and a report on the cross moderation exercise. A study based on the summer 2002 examinations*. Organised by the Assessment and Qualification Alliance on behalf of the Joint Council for General Qualifications.

Arlett, S. (2003). *A comparability study in VCE Health and Social Care units 3, 4 and 6. A review of the examination requirements and a report on the cross moderation exercise. A study based on the summer 2002 examinations*. Organised by the Assessment and Qualification Alliance on behalf of the Joint Council for General Qualifications.

Barry, K. (1997). An analysis of the relative demands of advanced GNVQ science and A-level Chemistry. *Journal of Further and Higher Education*, **21**, 1, 45–53.

Bloom, B.S. (Ed.) (1956). *Taxonomy of Educational Objectives – Book 1 – Cognitive Domain*. Michigan: Longman.

Coles, M. & Matthews, A. (1995). *Fitness for purpose: a means of comparing qualifications. A report to Sir Ron Dearing to be considered as part of his review of 16–19 education*.

Coles, M. & Matthews, A. (1998). *Comparing qualifications – Fitness for purpose. Methodology paper*. London: Qualifications and Curriculum Authority.

Crisp, V. & Novaković, N. (2009). Are all assessments equal? The comparability of demands of college-based assessments in a vocationally-related qualification. *Research in Post Compulsory Education*, **14**, 1, 1–18.

Edwards, E. & Adams, R. (2003). *A comparability study in GCE Advanced Level Geography including the Scottish Advanced Higher Grade Examinations. A review of the examination requirements and a report on the cross moderation exercise*. A study based on the Summer 2002 Examination and organised by WJEC on behalf of the Joint Council for General Qualifications.

Gagne, R.M. (1985). *The conditions of learning and theory of instruction* (4th Ed.). New York: Holt, Rinehart and Winston.

Guthrie, K. (2003). *A comparability study in GCE business studies units 4, 5, and 6 VCE business units 4, 5, and 6. A review of the examination requirements and a report on the cross moderation exercise*. A study based on the summer 2002 examinations. Organised by EdExcel on behalf of the Joint Council for General Qualifications.

Hughes, S., Pollitt, A. & Ahmed, A. (1998). *The development of a tool for gauging the demands of GCSE and A-level examination questions*. Paper presented at the British Educational Research Association Annual Conference, The Queen's University of Belfast.

Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen.

Learning and Skills Council (2009). Jargon Buster http://www.lsc.gov.uk/Jargonbuster/Vocational+certificate+of+education.htm [Accessed September 2009]

Mitchel, L. & Bartram, D. (1994). The place of knowledge and understanding in the development of National Vocational and Scottish Vocational Qualifications. In: *Competence & assessment briefing series no. 10*.

OCR (2009). http://www.ocr.org.uk/qualifications/type/nvq/index.html [Accessed January 2010]

OCR (2009). http://www.ocr.org.uk/qualifications/type/vrq/index.html [Accessed January 2010]

Ofqual (2008). Glossary http://www.ofqual.gov.uk/501.aspx#I [Accessed January 2010]

Pollitt, A., Ahmed, A. & Crisp, V. (2007). The demand of examination syllabuses and question papers, 166–206. In: P. Newton, J. Baird, H. Goldstein, H. Patrick & P. Tymms (Eds.) *Techniques for monitoring the comparability of examination standards*. London: QCA.

QCA (2006a). *Comparability study of assessment practice: Personal license holder qualifications*, QCA/06/2709 http://www.ofqual.gov.uk/files/personal_licence_holder_quals_comparability_study.pdf

QCA (2006b). *Comparability study of assessment practice Door supervision qualifications* QCA/06/2710 [on line] Available at: http://www.ofqual.gov.uk/files/door_supervision_quals_comparability_report.pdf [Accessed September 2009].

QCDA (undated). Glossary http://testsandexams.qcda.gov.uk/15862.aspx#S [Accessed January 2010].

Savory, C., Hodgson, A. & Spours, K. (2003). *The Advanced Vocational Certificate of Education (AVCE): A general or vocational qualification? Broadening the Advanced Level Curriculum*. IoE/Nuffield Series Number 7, School of Lifelong Education and International Development, Institute of Education, University of London [on line] http://www.ioe.ac.uk/schools/leid/nuff/rep7.pdf [Accessed September 2009]

SCAA (1995). *Report of a comparability exercise into GCE and GNVQ business*. London: School Curriculum and Assessment Authority.



# Developing and piloting a framework for the validation of A levels

**Stuart Shaw** CIE Research **and Victoria Crisp** Research Division

## Introduction

This article reports briefly on a current strand of research which aims to develop a methodology for validating general academic qualifications such as A levels. Validity is a key principle of assessment, a central aspect of which relates to whether the interpretations and uses of test scores are appropriate and meaningful (Kane, 2006). For this to be the case, various criteria must be achieved, such as good representation of intended constructs, and avoidance of construct-irrelevant variance. Additionally, some conceptualisations of validity include consideration of the consequences that may result from the assessment, such as affects on classroom practice. The kinds of evidence needed may vary depending on the intended uses of assessment outcomes. For example, if assessment results are designed to be used to inform decisions about future study or employment, it is important to ascertain that the qualification acts as suitable preparation for this study or employment, and to some extent predicts likely success.