

### Critical review

## Does assessing project work enhance the validity of qualifications? The case of GCSE coursework

by Victoria Crisp ([crisp.v@cambridgeassessment](mailto:crisp.v@cambridgeassessment))

### Contextualisation

Assessments other than written examinations are included in some qualifications in order to measure skills that are difficult to assess via exams, and hence enhance the validity of the qualification by better representing the skills relevant to that domain. The assessment of project work is one such method. GCSEs are the main qualifications taken by 16 year olds in England, Wales and Northern Ireland at the end of compulsory schooling. From their introduction, GCSEs in many subjects have included an assessment of project work, known as coursework, conducted by students in school and marked by their teachers. This paper takes a critical look at how the validity of a qualification might be enhanced by the inclusion of an assessment of school-based project work using GCSE coursework as a case study. Coursework is due to be removed from some GCSE subjects due to concerns about the burden placed on teachers and students and issues around authentication. This makes it very timely to discuss the validity of this assessment form. This literature review provides a backdrop for the author's doctoral studies at the Institute of Education, University of London, investigating the judgement processes that underpin coursework marking.

**Abstract:** *This paper begins by describing current views on validity and how certain assessment forms, such as school-based project work, may enhance validity. It then touches on debates about the dependability of assessment by teachers. GCSEs and GCSE coursework are then described along with the reasons for the inclusion of coursework in many GCSEs. Crooks, Kane and Cohen's (1996) chain model of eight linked stages of validity enquiry is then used as a structure within which to consider the validity of project work assessments, and specifically GCSE coursework assessment, drawing on the available literature. Strengths for validity include the ability to assess objectives that are difficult to test in written examinations, promoting additional skills such as critical thinking, creativity and independent thinking, and improving motivation. Possible threats to validity include the potential for internet and other types of plagiarism, tasks becoming overly structured and formulaic thus reducing the positive impact on learning, and the potentially heavy workload for teachers and students. The paper concludes by describing current policy changes in the UK with regard to GCSE coursework and relates this to strong and weak validity links for project work as a mode of assessment.*

### Introduction

A desire to encourage and assess skills not easily tested in written examinations underpins the use of project work assessments in a range of contexts. The general principle is that of increasing construct validity by testing relevant attributes. However, the use of such assessments can introduce other problematic factors. For example, Linn, Baker and Dunbar (1991) describe the use of portfolios in the Iowa school district. Teachers and students decide on what to put into the folders and on the amount of time perfecting their entries. Linn, Baker and Dunbar describe how this has the potential to reduce the breadth of a student's learning if too much emphasis is placed on a small number of pieces of work that will form their portfolio entry. This paper will discuss aspects of the validity of project work assessment, using GCSE coursework as an example.

## **Assessment validity**

Validity is a central concept in assessment. Broadly speaking, the validity of an assessment is about the degree to which it really measures what it purports to measure (Borsboom, Mellenbergh and van Heerden, 2004). More technically, validity is about the appropriateness of inferences that are made from assessment outcomes (Messick, 1989). The traditional view of validity is that there are different kinds of validity: content validity - how appropriate the content of the assessment is as a test of what it aims to assess; construct validity - how well the assessment measures appropriate underlying constructs; and criterion-related validity - how well an assessment relates to actual performance on a specified criterion; this can be predictive or concurrent. In the last few decades most validity theorists have come to consider the construct-content-criterion framework inadequate on the grounds that content and criterion-related validity are actually examples of evidence that support construct validity. Both Cronbach (1988; 1989) and Messick (1989) consider construct validity the central form. Within this view, Messick describes two main threats to construct validity: 'construct under-representation' - the assessment fails to capture important aspects of the construct, and 'construct-irrelevant variance' - capabilities that are irrelevant to the construct are assessed.

Around the same time there was also an emerging view that the concept of validity should be extended to include the consequences of assessments (Cronbach, 1988; Messick, 1989; Shepard, 1993) such as the use of test results, impact on instruction and social consequences. In the climate of both these revisions to the dominant notion of validity, attempts have been made to characterise the types of evidence needed to support construct validity (eg, Frederiksen and Collins, 1989; Messick, 1989; Linn, Baker and Dunbar, 1991; Crooks, Kane and Cohen, 1996).

Reliability is about whether an assessment is repeatable and measures consistently, with a minimum of error. Much attention is given to this issue in assessment development and procedures. Contemporary thought considers reliability a constituent part of construct validity (eg, Frederiksen and Collins, 1989, Crooks, Kane and Cohen, 1996). Crooks, Kane and Cohen (1996) see inter-marker and intra-marker reliability as part of validity because they affect the confidence with which inferences can be made. In the case of school-assessed project work, the intention for its use is to improve validity but it may mean greater risks for reliability. Risks to reliability can be minimised by quality control procedures.

Written examinations can arguably only test certain types of skills. There are strong arguments for authentic assessments to test vocational skills, practical assessments to assess practical skills and project work to assess investigative skills (see for example, Frederiksen and Collins, 1989; Linn, Baker and Dunbar, 1991; Murphy *et al*, 1995; Nuttall, 1990; Tolley *et al*, 2003). In many cases, depending on the skills important to that domain, a qualification using a range of assessment types can theoretically better reflect the skills important to the domain and hence should enhance validity by providing better coverage of relevant constructs.

## **Teacher marking**

There is continuing debate over whether teachers can be trusted to assess their own students for summative purposes. Some argue that teachers' judgements cannot be free from bias whilst others claim that assessment by teachers is the most valid method, as they see a student's work over a period of time, and that teachers' professional judgements should be trusted. Research evidence shows that the validity and reliability of teacher assessment varies and may be related to certain features such as the degree of specification of tasks and criteria (Harlen, 2004), school cultures (Ellis, 1998) and moderation procedures.

Moderation procedures can be used to help ensure the validity of teacher assessments. Moderation usually involves an external examiner checking a sample of a teacher's or school's marking to ensure it is in line with external standards and with marking in other schools. Such procedures can help to reduce any effects of teacher bias by standardising judgements, and thus can help to provide valid and reliable measures of student achievement based on teacher marking.

In some countries teacher assessment plays a much larger role than in the UK. For example, in Sweden, assessment and grading is conducted by teachers and is seen as an exercise of teacher authority in school law about which students had no right of appeal until fairly recently (Andersson, 2004). Assessments must be conducted in line with national goals and grading criteria. However, teachers have to interpret these criteria for themselves and balance pressures to provide equity and justice against pressures to produce acceptable results in terms of not making themselves or the school 'look bad' (Andersson, 2004). Despite some of the issues it seems that assessment by teachers is more publicly accepted in Sweden than in the UK where the outcomes of national assessments via centralised tests have been more widely used as indicators of school performance than the accompanying assessments made by teachers. There is also a much greater respect for teacher judgement in Australia where all states use school-based assessment to some extent in reporting at the end of secondary schooling and in Queensland final assessments are based completely on peer-moderated school assessments (Cumming and Maxwell, 2004).

### **GCSE coursework**

The GCSE (General Certificate of Secondary Education) was introduced in 1988 (first awarding) and is available in a wide range of subjects. It is the main qualification taken by 16 year olds in England, Wales and Northern Ireland at the end of compulsory schooling. GCSEs are assessed mainly by traditional examinations but in many subjects a percentage of the assessment is via coursework. In the National Criteria Glossary of Terms coursework is defined as "all types of activity carried out by candidates during their course of study and assessed for examination purposes" (SEC, 1986, p 1). Coursework takes a wide range of forms across subjects: from written reports of fieldwork in geography to musical performances, and from pieces of art work to oral contributions during lessons in English. Coursework tends to involve pieces of student work produced over a period of time and usually completed specifically with the intention that the particular piece of work be submitted for the coursework assessment, in contrast to it involving the post-hoc selection of exemplar work that students undertake during their learning experiences. GCSE coursework, in most cases, is assessed by teachers, internally moderated across teachers within schools and then externally moderated by examiners.

Evaluating project work was not a new method of assessment: for example, it was an optional element of the precursor qualifications, the Certificate in Secondary Education or CSE. However, it was the introduction of GCSE that saw a much increased presence of project work as part of the assessment culture in the UK. Coursework was included in many GCSEs to increase the validity of assessment by providing wider evidence of student work and to enhance pupil learning by valuing skills such as critical thinking and independent learning (SEC, 1985). As the Secondary Examinations Council (SEC, 1985, p 6) put it:

above all, the assessment of coursework can correspond much more closely to the scale of values in this wider world, where the individual is judged as much by his or her style of working and ability to cooperate with colleagues as by the eventual product.

The validity, and its subcomponent reliability, of GCSE coursework has come under much discussion since its introduction with the focus of concerns changing over time. At the

inception of the GCSE the main threats anticipated were possible unreliability of teacher marking, possible cheating and concern that girls would be favoured (see QCA, 2006). Now, concerns about consistency across similar subjects, authenticity of student work including the issues of internet plagiarism and excessive assistance from others, tasks becoming overly-structured and hence reducing learning benefits, along with the overall burden on students across subjects, have become prominent.

In 2005, the 14-19 Education and Skills White Paper (DfES, 2005) gave QCA (Qualifications and Curriculum Authority) a remit to reconsider the value of coursework and address possible concerns. The initial review (QCA, 2005) involved questionnaires to centres, interviews with teachers, candidates and parents, statistical research and a conference day with examiners. Further work has included a MORI telephone survey of teachers' views (MORI, 2006) and a review using this evidence and evidence from QCA monitoring (QCA, 2006).

### **The validity of project work assessment and, specifically, GCSE coursework**

The work of Crooks, Kane and Cohen will be used to provide a structure within which to discuss the potential validity enhancements and threats associated with the inclusion of project work assessments in qualifications, with specific reference to GCSE coursework assessment. Crooks, Kane and Cohen's set of criteria for evaluating validity has been chosen over other frameworks as it allows us to focus on the validity that assessments of project work may add as part of a full qualification, in comparison to qualifications based only on written examinations. In addition, some other sets of criteria have been criticised or have certain disadvantages. For example, the framework proposed by Frederiksen and Collins (1989) focuses on performance assessment and structures the consideration of validity around the following categories: transparency, directness of measurement, scope, and reliability. Whilst these areas address important components of validity particularly for performance assessments, this framework has been criticised by Messick (1994) for its limited nature which may result in certain important aspects of validity, such as score interpretation, being overlooked. Crooks, Kane and Cohen's criteria framework is aligned with Cronbach's (1988) perspectives on validity which has strengths in terms of its consideration of the social consequences of assessment. Crooks, Kane and Cohen transform Cronbach's ideas into a tool to direct practical consideration of validity. It also maps onto Messick's (1995, p 11) "aspects of construct validity" but again provides a more a practical scaffold for evaluating validity. Additionally, there is substantial overlap between the structure suggested by Crooks, Kane and Cohen and some other suggested criteria (eg, Linn, Baker and Dunbar, 1991). Crooks, Kane and Cohen (1996) depict assessment validity enquiries as a chain of eight linked stages in order to provide a structure for considering the validity of an assessment. The stages defined are: administration, scoring, aggregation, generalisation, extrapolation, evaluation, decision and impact. For each stage possible threats to validity are exemplified. Crooks, Kane and Cohen suggest that considering possible threats at each stage will allow any 'weak links' to be identified for an assessment.

#### ***Administration of the assessment tasks to the student***

The conditions under which students take an assessment can impact on the validity of interpretations about the assessment and this link in Crooks, Kane and Cohen's model involves examining the task administration. The use of project work eases the threat to validity caused by stress in exams and is thought to improve motivation. For example, coursework is thought to be fairer for hard-working pupils who are affected by exam stress and also allows the use of tasks that would cause anxiety in an exam situation (SEC, 1985; 1986). However, the testing conditions involved in coursework can be dissimilar (Scott, 1990) and clashing deadlines for coursework completion across subjects may cause anxiety for some students.

The threat to validity that currently seems to be considered most significant for GCSE coursework comes under the category of 'administration' and is about ensuring authenticity of student work. Concerns relate to plagiarism and excessive assistance from others. The arrival of the internet and increased presence of computers in homes has made the potential for plagiarism greater. Additionally, the level of structure and uniformity of coursework tasks may make plagiarism easier.

A limited degree of engagement of parents in their child's GCSE coursework is encouraged eg, encouraging their child to spend time on their coursework and providing access to resources. QCA (QCA, 2005) found that nearly two-thirds (63 per cent) of parents helped in some way, for example by checking spelling and grammar, and five per cent of parents admitted to actually drafting some of their child's coursework. Collusion was always a possibility with coursework but seems to be greater concern now than in the past.

The QCA review (2005) reports that some students admitted trying to download assignments from the internet but not to using them. Some admitted having submitted the work of a sibling or friend as their own. There is also a possibility for inadvertent collusion between peers where part of fieldwork or investigations involves group work or identical tasks.

The QCA (2005) report makes a number of proposals including the development of guidelines for teachers and parents on the limits of permitted help and giving a higher profile to malpractice. Guidelines have now been produced and made available and these measures may help to reduce potential threats to validity in this link.

### ***Scoring of the student's performances on the tasks***

With regard to the scoring of an assessment, Crooks, Kane and Cohen propose that aspects that could reduce the validity of score interpretations and consequent decisions are considered. If scoring failed to capture important qualities of task performance for an assessment this would be a threat to its validity. Additionally, if scoring criteria are made too objective in an attempt to increase rater agreement, this could potentially narrow the factors included in scores. Coursework assessment offers an improvement on written examinations in that there is less risk of scoring emphasising unimportant but easily rated aspects of student performance. However, whilst it has been argued that providing wider evidence of pupil work through coursework will increase the repeatability of the assessment (SEC, 1985; 1986), it was always acknowledged that monitoring the marking reliability associated with GCSE coursework assessment would be important. Indeed, many of the negative responses to the introduction of GCSE involved fears that coursework marking would be unreliable and easily open to abuse (Kingdon and Stobart, 1988). Leonard (1991) discusses the "tension between trusting the professional judgement of teachers and the issue of public confidence in the system of assessment" (p 10). It is perhaps counter-intuitive to public opinion that teachers can judge their own students without bias.

Some data are available on the reliability of coursework marking. Taylor (1992) asked two moderators to re-mark pieces of coursework in each of GCSE English, maths and history and A-Level psychology and compared the marks given between the two moderators with the mark given by the original moderator. Good correlations between different pairs of moderators were found in each subject (ranging from 0.73 to 0.97). Wiliam (1996) mentions evidence that in the marking of the 100 per cent coursework English GCSE teachers learnt to agree on what grade a piece of coursework was worth but they did not always agree on the aspects of the work that were most significant in making the work worth a particular grade.

Crooks, Kane and Cohen (1996, p 272) comment that "it is desirable to reduce the extent of such inconsistency, but not at the expense of eliminating or reducing the weight given to important aspects of task performance which can only be assessed through professional

judgement". This view would support the inclusion of project work components in qualifications where they are likely to facilitate the assessment of relevant skills, even at the expense of slightly reduced marking consistency.

### ***Aggregation of the scores on individual tasks to produce one or more combined scores (total score of subscale scores)***

Issues under Crooks, Kane and Cohen's aggregation link include aggregating tasks that are too diverse and giving inappropriate weights to different aspects of assessment. Whilst the aggregation of scores from coursework and other examined components to determine GCSE grades could be considered an aggregation of diverse tasks, this is not generally considered a problem for the use of coursework.

### ***Generalisation from the particular tasks included in a combined score to the whole domain of similar tasks (the assessed domain)***

This link considers the accuracy of generalising from an aggregated score in an assessment to performance in the assessed domain, i.e. the entire range of tasks falling within the specification. If the conditions of the assessment vary too much then this would make such generalisations problematic. The term 'reliability' would often be used to describe this issue. With coursework, the conditions do vary somewhat and the tasks used vary but this may be necessary in order for coursework to achieve its purpose of broadening the skills assessed without becoming so over-defined that the learning benefits are lost and risks of plagiarism are increased. If the sample of each student's work assessed is small, this could be a threat to reliability. Coursework tends to involve just one or two tasks but these are large tasks conducted over a period of time so they effectively increase the sample size for a GCSE qualification more than could be achieved using an equivalent exam, and hence should help to avoid 'construct under-representation' (Messick, 1989).

### ***Extrapolation from the assessed domain to a target domain containing all tasks relevant to the proposed interpretation***

In the extrapolation link we consider the validity of extrapolating assessment results from the assessed domain to the target domain. This might usually be termed 'generalisability'. According to Crooks, Kane and Cohen, overly constrained assessment conditions would threaten validity in the extrapolation link. This threat to validity is likely to be reduced by the inclusion of a project work assessment as part of a qualification.

Another potential threat to validity in terms of extrapolation occurs if parts of the target domain are not assessed or are given minimal weight. This is similar to 'construct under-representation' as described by Messick (1989). The inclusion of coursework in GCSE assessment is likely to improve validity in this respect as it allows types of skills that cannot be assessed by an examination to be evaluated. As already described, improving construct representation was one of the key aims of including coursework in GCSE from the outset.

### ***Evaluation of the student's performance, forming judgements***

This link in the assessment chain is about evaluating what the scores relating to the target domain mean, for example, evaluating what the scores tell us about a student's strengths and weaknesses. Potential threats to validity at this stage can include biased interpretations of test scores, for example, as a result of a 'halo effect', and poor understanding of the assessment information and its limitations. These issues are the same for GCSE results

regardless of whether coursework formed part of the assessment and are hence beyond the scope of the current discussion.

### ***Decision on actions to be taken in light of judgements***

Crooks, Kane and Cohen's 'decision' link is about actions that are taken as a result of judgements, for example, decisions regarding whether a student's score is considered appropriate to admit them to a course. When evaluating the validity of an assessment this stage involves considering the merit of the decisions that are taken and whether they are consistent with the information on which they are based. One possible threat to validity at this stage according to Crooks, Kane and Cohen would be poor pedagogical decisions. The inclusion of project work assessments should give space for teachers to make good pedagogical decisions. They have more scope to provide useful feedback to students and greater freedom and flexibility within the curriculum. The latter was reported by teachers in MORI's survey about GCSE coursework for QCA (MORI, 2006). However, there is a risk that some teachers might dedicate too much time to coursework at the expense of other areas of study.

### ***Impact on the student and other participants arising from the assessment processes, interpretations, and decisions***

This link in the assessment process, as described by Crooks, Kane and Cohen, looks at consequential aspects of validity. The direct and indirect impacts of assessment are to be considered along with the effects of experiencing the whole process of assessment. Crooks, Kane and Cohen suggest that threats to validity here can include positive consequences not being achieved or the occurrence of a negative impact of the assessment. The inclusion of coursework in the GCSE was intended to have a positive impact on validity in this respect by providing a number of benefits to learning such as promoting skills of critical thinking, creativity, independent thinking, communication, research and reflection on work (SEC, 1985; 1986) and allowing helpful feedback from teachers (Ogborn, 1991). Coursework was also intended to be motivating through the realistic sense of audience, the opportunity to negotiate the task, and continual assessment (SEC, 1985; 1986; Ogborn, 1991). In addition, Ogborn (1991) argues that coursework forces teachers to plan courses carefully. In these ways, the use of coursework might reduce some potential threats to validity to do with impact compared to cases where assessment consists of examinations alone, which may encourage a focus on factual knowledge at the expense of higher order and practical skills. However, if concerns about coursework becoming overly formulaic and predictable in some subjects are well-founded, then coursework may not achieve its intended positive impact in those areas. Achieving positive consequences may also be at risk if some students only engage with coursework tasks at a surface level.

Additionally, the heavy workload for teachers and students that is sometimes reported would constitute a negative impact of coursework for some of those involved and hence may weaken validity in this respect. In the early days of the GCSE, efforts were made to address this concern and various teachers and Local Education Authority professionals investigated and sought to provide advice and good practice ideas based on experience. The main means of controlling the demand of coursework is thought to be to "ensure that coursework is integrated into the curriculum" (SEC, 1985, p 8) with tasks arising out of good classroom practice (Cunningham, 1991). Possibilities such as using one piece of coursework to address requirements of more than one subject (Leonard, 1991) or to use methods other than writing were tested but did not become common practice, except for the current overlap in coursework between English and English Literature. Cross-curricular schemes required extra planning from teachers but did reduce student workloads (Leonard, 1991).

It is interesting to note that Scott (1990) found that only a small number of pupils were doing excessive amounts of coursework and other homework. He also reported that the way that pupils reacted to coursework and homework pressure was not related to the amount they actually had to do.

## **Strong and weak validity links for including project work assessment**

Using Crooks, Kane and Cohen's model we can identify the links where project work assessment reduces threats to validity compared with examinations alone and links where threats to validity remain. Project work, and specifically coursework, has particular strengths in terms of improving construct representation (extrapolation), the potential for positive effects on learning (impact) and increasing motivation and reducing assessment anxiety (administration). When GCSEs began, the threats to validity that caused concern were possible negative effects due to workload for teachers (impact) and the potential for biased or inconsistent marking by teachers (scoring). Recently, concerns have shifted towards the issue of authenticating work (administration) and it is this threat to validity, combined with workload issues for students and teachers that seem to be central in driving current changes. Most of the issues raised here for GCSE coursework seem likely to apply to other summative assessments of project work.

## **The future of GCSE coursework**

The 2005 QCA report concluded that the use of coursework needs review in a number of subjects and that it may not be needed in some subjects. A series of reviews were instigated starting with one focused on mathematics and a MORI study of teachers' views across seven subjects (QCA, 2006). The QCA confirmed that coursework would be dropped from GCSE mathematics from courses beginning in September 2007 and from a number of other subjects (business studies, classical subjects, economics, English literature, geography, history, modern foreign languages, religious studies and social sciences) from courses beginning in September 2008 where they will be replaced with controlled assessments. Controlled assessments are likely to involve tasks being set or approved by the awarding body, conducted under supervised conditions and marked by teachers (QCA, 2007). This would mean a reduction in possible threats to validity in terms of authentication (administration link) and perhaps in terms of marking reliability (scoring link). However, it could have the potential to reduce the validity benefits of coursework in terms of construct representation (extrapolation link) if tasks limit the skills tested, or to reduce validity benefits in terms of impact if tasks become less interesting and motivating or overly structured.

Decisions over changes have been justified by QCA on the basis of three key principles: that the intended learning outcomes in the subject should be critical in determining the appropriate form of assessment, that the most valid (including reliable) form of assessment for a learning outcome should be used so that results are fair and robust and maintain confidence, and that the assessment process should be manageable (QCA, 2006). It is interesting that the teachers interviewed by MORI (2006) were fairly positive about coursework, and the QCA reviews report a general consensus view that the benefits outweigh the drawbacks (QCA, 2005). Concerns about internet plagiarism were not as great as might have been expected (82 per cent of teachers disagreed that students used the internet too much) and whilst more than half of teachers felt that students in some schools can gain unfair advantage in the current system, the most frequently mentioned drawback was the burden of marking coursework. The interviews by MORI found that 66 per cent of teachers were opposed to removing coursework and 51 per cent were strongly opposed to its removal. The MORI interview evidence would not seem to support the decisions that have been made. However, the replacement of coursework with 'controlled assessments' rather

than exams, may address teacher concerns that removing coursework would impact on teaching and learning whilst allowing a continued focus on wider skills.

It seems that concerns about threats to validity in the administration link, i.e. authenticity and burden seem to be out-weighing possible advantages of coursework to validity in terms of construct representation (extrapolation link) and learning experiences (impact link). However, if the controlled assessments could maintain validity in terms of construct representation and learning experiences, as well as reducing threats in relation to administration, then they could potentially provide a more robust overall 'chain' of validity links.

## Discussion

The issues around GCSE coursework and the currently unfolding policy changes provide a case study within which to consider the use and validity of teacher-assessed project work as an assessment form. It seems that theoretically project work should allow a broader and more relevant set of skills to be tested but that issues around authenticating student work threaten public confidence in the UK context. It is interesting that the new diploma qualifications currently being introduced in the UK involve a greater use of teacher-assessed project work. This is motivated by a desire to promote the development of skills such as communication, planning, time management and research which are relevant to both further education and employment and indicates a clear acknowledgement of this likely benefit to validity. However, project-based assessments within the diplomas will be 'controlled' similarly to the coursework replacement assessments. This suggests that similar concerns regarding authenticity and public confidence have also influenced the choice of assessment mode. This paper has analysed the validity of project work as a mode of summative assessment. It suggests clear benefits to this assessment type but that achieving sufficient control is important.

## References

- Andersson, H. (2004) "*Seeking high and low*": teachers' grading practice in time-table free schools in Sweden. Paper presented at the European Conference on Educational Research, Crete, September 2004.
- Borsboom, D., Mellenbergh, G. and van Heerden, J. (2004) The concept of validity. *Psychological Review*, 111, 4, 1061-1071.
- Cronbach, L. (1988) Five perspectives on validity argument, in H. Wainer and H.I. Braun (Eds) *Test validity* (pp 3-17). Hillsdale, NJ: Erlbaum.
- Cronbach, L. (1989) Construct validity after thirty years. In R. Linn (Ed.) *Intelligence: measurement, the theory and public policy* (pp 147-171). Urbana: University of Illinois Press.
- Crooks, T., Kane, M. and Cohen, A. (1996) Threats to the valid use of assessments. *Assessment in Education: Principles, Policy and Practice*, 3, 3, 265-286.
- Cumming, J. and Maxwell, G. (2004) Assessment in Australian schools: current practice and trends. *Assessment in Education: Principles, Policy and Practice*, 11, 1, 89-108.
- Cunningham, B. (1991) Coursework in GCSE English and English literature. In SEAC, *Coursework: learning from GCSE experience: an account of the proceedings of a SEAC conference*. London: Secondary Examinations and Assessment Council.

- DfES (2005) *14-19 Education and Skills White Paper*. London: DfES.
- Ellis, S. (1998) *Developing whole-school approaches to curriculum and assessment in secondary schools*. Paper presented at the British Educational Research Association Annual Conference, Queen's University, Belfast.
- Frederiksen, J. and Collins, A. (1989) A systems approach to educational testing. *Educational Researcher*, 18, 9, 27-32.
- Harlen, W. (2004) A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes, in *Research Evidence in Education Library*. London: EPPI – Centre, Social Sciences Research Unit, Institute of Education.
- Kingdon, M. and Stobart, G. (1988) *GCSE examined*. Lewes: Falmer Press.
- Leonard, A. (1991) Case studies of school-based innovation. In: SEAC, *Coursework: learning from GCSE experience: an account of the proceedings of a SEAC conference*. London: Secondary Examinations and Assessment Council.
- Linn, R., Baker, E. and Dunbar, S. (1991) Complex, performance-based assessment: expectations and validation criteria. *Educational Researcher*, 20, 8, 15-21.
- Messick, S. (1989) Validity, in R. Linn (Ed.) *Educational measurement* (3<sup>rd</sup> ed.) (pp 13-103). New York: American Council on Education.
- Messick, S. (1994) The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 2, 13-23.
- Messick, S. (1995) Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 9, 741-749.
- MORI (2006) *Teachers' views on GCSE coursework: research study conducted for the QCA*. Available at:  
[http://ofqual.gov.uk/files/QCA-06-2735\\_MORI\\_report\\_teacher\\_views\\_coursework.pdf](http://ofqual.gov.uk/files/QCA-06-2735_MORI_report_teacher_views_coursework.pdf)  
[accessed 15 September 2008].
- Murphy, R., Burke, P., Content, S., Frearson, M., Gillespie, J., Hadfield, M., Rainbow, R., Wallis, J. and Wilmut, J. (1995) *The reliability of assessment of NVQs*. Report to the National Council for Vocational Qualifications, School of Education, University of Nottingham.
- Nuttall, D. (1990) The validity of assessments (appendix II), in A. Wolf and R. Silver (Eds) *Work based learning: trainee assessment by supervisors*. Manpower Services Commission, Research & Development: No. 33.
- Ogborn, J. (1991) In-course assessment: creating and making best use of opportunities, in *Coursework: learning from GCSE experience: an account of the proceedings of a SEAC conference*. London: Secondary Examinations and Assessment Council.
- QCA (2005) *A review of GCE and GCSE coursework arrangements*. London: Qualifications and Curriculum Authority. Available at: <http://www.qca.org.uk/libraryAssets/media/qca-05-1845-coursework-report.pdf> [accessed 15 September 2008].

- QCA (2006) *A review of GCSE coursework*. London: Qualifications and Curriculum Authority. Available at: <http://image.guardian.co.uk/sys-files/Education/documents/2006/10/06/courseworkreport.pdf> [accessed 15 September 2008].
- QCA (2007) *Controlled Assessments*. London: Qualifications and Curriculum Authority. Available at: [http://ofqual.gov.uk/files/qca-07-3208\\_controlledassessmentfinal3.pdf](http://ofqual.gov.uk/files/qca-07-3208_controlledassessmentfinal3.pdf) [accessed 15 September 2008].
- Scott, D. (1990) *Coursework and coursework assessment in the GCSE*. Coventry: University of Warwick, Centre for Educational Development, Appraisal and Research, CEDAR reports, no.6.
- SEC (1985) *Working paper 2: coursework assessment in GCSE*. London: Secondary Examinations Council.
- SEC (1986) *Working paper 3: policy and practice in school-based assessment*. London: Secondary Examinations Council.
- Shepard, L. (1993) Evaluating test validity. *Review of Research in Education*, 19, 405-450.
- Taylor, M. (1992) *The reliability of judgements made by coursework assessors*. AEB Research Report RAC 577.
- Tolley, H., Greatbatch, D., Bolton, J. and Warmington, P. (2003) *Improving occupational learning: the validity and transferability of NVQs in the workplace*. Sheffield: DfES.
- William, D. (1996) Standards in examinations: a matter of trust? *The Curriculum Journal*, 7, 3, 293-306.