JEM Journal of

EDUCATIONAL MANAGEMENT

A Bi-annual Publication of
THE INSTITUTE FOR EDUCATIONAL PLANNING AND ADMINISTRATION
(IEPA)
University of Cape Coast, Ghana.

ISSN 0855-3343 Vol. 3 Nov. 2000

Journal of Educational
Management Vol.3 (110-127)
November 2000

ISSUES IN COMBINING CONTINUOUS ASSESSMENT SCORES WITH EXTERNAL EXAMINATION SCORES FOR CERTIFICATION AT THE BASIC AND SECONDARY SCHOOL LEVELS

F. K. Amedahe

Abstract

Combining continuous assessment scores with external examination scores for the certification and selection of junior and senior secondary school graduates has gained currency in some countries, including Ghana and Nigeria. While the rationale behind combining continuous assessment scores with external examination scores appears to be sound, there are some fundamental issues which must be understood and addressed by policy makers and practitioners in order to make informed decisions concerning the practice. This paper addresses some relevant issues relating to the policy and practice of combining the two sets of scores for certification. The issues include differences in the quality of teacher assessments and external examinations, methods of moderating teacher assessment scores before combining them with external examination scores and challenges of moderating teacher assessment scores. Some suggestions are made concerning how some of the issues may be resolved.

Introduction

The last two decades can be described as a period of rejuvenation and transformation of the field of educational measurement with emphasis on assessment. This period has been characterized by the development of new measurement theories such as the item response theory, the reconceptualization of the concept of validity in testing, and the search for alternative ways of assessing student achievement to obtain more valid results. During the period, some countries including Ghana and Nigeria, for example, in their search for better ways of assessing and certifying students, introduced the concept and practice of combining teacher-based continuous assessment scores and external examination scores at the basic and secondary school levels into their education systems. For instance, in Ghana the

implementation of the New Structure and Content of Education in the late 1980s and early 1990s included implementing the practice of combining teacher-based continuous assessment scores with external examination scores for certifying junior and senior secondary school graduates. The assessment policy framework within the reform package stipulated that the final grading of junior and senior secondary school students should be made up of 40% continuous assessment marks and 60% external examination scores (Ministry of Education, Subcommittee Report, 1988; Opong, 1992). This weighting of the assessments was recently changed to 30% and 70%, respectively. Similarly, in the first Senior School Certificate Examination held in Nigeria in May/June 1988, teacher-based continuous assessment scores were combined with external examination scores to certify graduates (Awuwoloye, 1988). In the West African sub-region, particularly in the Anglophone countries, external examinations for certifying graduates at the basic and secondary levels of education are solely conducted by the West African Examinations Council (WAEC) on behalf of the education ministries of the respective countries.

The rationale behind using teacher-based continuous assessment as a component score for certifying graduates is that it is difficult, if not impossible, to fully capture all that a student has achieved in several years of study in a programme in a one-shot test lasting between one and three hours. It is thought that by assessing students continuously throughout a programme (by teachers), a better picture of the student's achievement would emerge. However, teacher assessment alone cannot provide the necessary and sufficient information on students' attainment in terms of the same standard across schools in a country because of the non-standardisation of teacher assessments; hence the need for an external standardised examination for students.

The use of different assessment methods, as the combination implies, agrees with *Principles for Fair Student Assessment Practices* (Joint Advisory Committee, 1993) in Canada and *Standards for Teachers' Competence in Educational Assessment of Students* (AFT, NCME, and NEA, 1990) in the United States. These professional documents on assessment indicate that different assessment modes and methods should be employed in assessing

students to ensure the collection of comprehensive and consistent information on them. This significant element is, nevertheless, missing in the use of a one-shot external examination alone in grading graduates. Some people also view the practice of continuous assessment as a way of providing timely feedback to the teacher and student for instructional and learning purposes.

Even though the rationale behind the practice of combining continuous assessment scores with external examination scores for certifying students is sound, there are some fundamental issues and concerns which must be clearly understood by policy makers and practitioners in order to make informed decisions regarding the use of the procedure. The informed decisions can aid in the form of modifying the practice to improve it, if it is so desired or discontinuing the practice. This paper addresses some issues considered relevant to the practice of combining teacher-based assessment scores with external examination scores from the theoretical and practical points of view.

Issues

The practice of combining teacher assessment scores with external examination scores is based on the premise that both teacher assessments and external examinations measure students' achievement in different subject areas in the same way. In other words, the practice assumes that teacher assessment scores and external examination scores are equally valid and reliable. This premise is not, in most cases, necessarily true. This has led to the practice of moderating teacher assessment scores before combining them with external examination scores for certifying graduates. The practice of combining teacher assessment scores with external examination scores begs some questions. These include: How comparable are teacher assessments with external examinations in terms of their quality? How comparable are teacher assessment results from one school to another? What methods can be used to moderate teacher assessment scores before combining them with external examination scores? How fair is moderating teacher assessment scores to all students? What weighting should be given to the components in the final grading of students? How can the validity of the scores resulting from the combination of the components be ascertained? What validity evidence of such composite scores

is available? Answers to the above questions and others are relevant in the implementation of the practice of combining teacher assessment scores with external examination scores for certifying graduates as a way of improving the practice. The rest of the paper is divided into four main sections. The first two sections address two main issues identified using theoretical and practical basis. The third section discusses the issues while the fourth section concludes the paper.

Quality of Teacher Assessments and External Examinations

The first major issue in combining teacher assessment and external examination scores is concerned with the quality of both assessments. The assessments and the examinations taken by the student are the building blocks of the final combined (composite) score obtained by a student. It follows, therefore, that the quality of the final results in terms of their validity and reliability will be determined by the quality of the prior assessments and the examination. Thus, the quality of teacher assessments as well as the external examination is crucial in determining the quality of the final composite score, assuming the procedure used for combining the scores is also valid.

The literature on the quality of teacher assessments is rather sparse. In the United States, for example, Stiggins and Conklin (1992) report that "research on testing in the schools has provided very little information concerning the quality of teacher developed assessments" (p.17). Studies by Fleming and Chambers (1983), Carter (1984), Gullickson and Ellwein (1985) indicate problems with the quality of teacher assessments. The studies show the need for teachers to write better items and longer tests, as well as value the use of statistical analysis of items as a helpful strategy in the classroom. These findings could be generalized to teacher assessment in developing countries, including Ghana and Nigeria. Despite the paucity of evidence of quality of teacher assessment, they are believed to be valid because they can be based on a long-term knowledge of each student within different contexts which external examinations lack (Harlen, 1994). Compared to external examinations, teacher assessments, perhaps, are more finely tuned to the content of their own teaching. Teacher assessments also tend to emphasize low order thinking skills such as recall and comprehension rather than high order skills.

With regard to reliability, teacher assessments are not considered to be highly reliable. Frisbie (1988) reported that the reliability of teacher-made tests is around 0.5 while that of external examinations is about 0.9. The validity claimed for teacher assessments, if any, is therefore obtained at the expense of low reliability because of an inherent tension between validity and reliability – as validity of test scores increases, the reliability of the scores decreases. This raises the question of whether teacher assessment can have a high validity in the presence of low reliability. A recent study using the classical test theory model by Feldt (1997) demonstrated that validity could rise when reliability declines, provided that irrelevant items are removed from the composite and relevant items remain or are added.

The issue of reliability of teacher assessments is crucial if they are to be used in certification of students. Certification of students demands high level of reliability in terms of consistency of performance and scoring, for comparability purpose.

In terms of quality, therefore, available evidence indicates that significant differences exist between teacher assessment and externally developed and administered examinations. Generally speaking, teacher assessments are known to be relatively low in quality, particularly in terms of their reliability, when compared to external examinations. This has led to high public scepticism about the use of teacher-based scores for high-stakes decisions including certification and selection. The relatively poor quality of teacher assessments stems from the fact that (1) for teachers, an assessment's quality is not primarily a technical matter but the assessment's impact on the teaching-learning process, (2) there is lack of external review of the nature and quality of teachers' assessments and this makes way for varying teacher standards applied to students, (3) more often than not, teachers do not receive adequate professional training in assessment techniques, and (4) teacher assessments rarely follow recommended practices of item development and refinement before administration (Amedahe, 1989, Stiggins, 1994, Stiggins & Conklin, 1992).

The crux of the quality issue is that teacher assessments tend to have some inherent weaknesses, particularly their non-comparability, hence the relatively low reliability of teacher-made assessment scores. As a result, combining teacher assessment

scores with external examination scores for certifying students is problematic. This does not mean that external examinations are perfect. No examination or assessment can yield perfect scores. Rather, external examinations are in most cases developed by educational assessment/ measurement and curriculum experts. As such, they are more carefully constructed and refined before administration. In addition they are, in most cases, administered under standardised conditions.

External examinations, on the other hand, are one-shot in characteristic that affects their validity. For instance, a candidate's performance in external examinations is subject to some extraneous factors including the psychological condition at the time of taking the examination as well as the physical condition in which the examination is administered. The one-shot nature of external examinations also limits the mode and the number of items that the student can be tested on. Sometimes, it happens that some candidates prepare well for some examinations but luck may not be on their side when what they emphasized in their learning is not well represented in the examination – a sampling issue. Under such a circumstance, the student's score may not be a true representation of her/his ability – a validity issue.

In any case, the quality of teacher assessments necessitates the moderation of the scores with the view to improving their comparability and/or reliability before they are combined with external examination scores. Moderating teacher assessment scores with its related issues are discussed next.

Moderation of Teacher Assessment Scores

Essentially, moderation refers to making scores from different schools and/or teachers comparable. Even though there are some variations in the use of the concept, Harris (1986) points out that comparability is the key element in moderation. The variation includes the use of the concept in some countries to mean the process of computing total examination scores for students taking different examinations. In this paper, moderation refers only to making scores from different schools and/or teachers comparable. It is noteworthy that in the United Kingdom (U.K.) and in some other countries, the terms quality assurance and quality control are now used in place of moderation depending on whether

the focus is on improving the process of assessment (quality assurance) or on only ensuring that the assessment outcome is judged in a comparable way–quality control (Gipps, 1994; Harlen, 1994).

It is important at the outset to note that the issue of moderating teacher assessment scores for certification and selection purposes by external examination agencies does not mean equating them with the examination scores. This is so because the conditions necessary for equating, namely equity, population invariance, symmetry of the two tests, and the two tests measuring the same ability (Lord, 1980; Kolen & Brennan, 1995; Petersen, Kolen & Hoover, (1989) are not met. Neither is the process a prediction or the usual regression issue. Limitations and inappropriateness of using regression procedure in making scores comparable includes bias and lack of symmetry of the scores as pointed out by Angoff (1979). However, a final composite score can be used to predict a student's future performance on a criterion.

A critical review of the literature on making different teachers' assessment (internal assessment) scores comparable indicates that several methods are used (Angoff, 1979; Bardell, Forrest and Shoesmith, 1978; Burton and Linn, 1993; Cohen and Deale, 1977; 1986; Harlen, 1994; Linn, 1993; Smith, 1978). In this paper, for simplicity, the methods are categorized into three: (1) moderation by inspection, (2) statistical moderation, and (3) moderation by monitoring. These are, generally, the most common methods of moderation. It is worthy to note that even though the methods are discussed under the above named categories, the lines between some of them are blurred.

Moderation by Inspection

Moderation by inspection involves bringing in persons, called moderators, to review, to re-grade, or to independently grade either a randomly selected sample or all students' responses to items on teacher assessment, particularly constructed responses (Burton and Linn, 1993; Cohen and Deale, 1977; Harlen, 1994). The goal is to determine whether a teacher's scores are out of line according to established standards or criteria. This process can result in ratification or repudiation of the teacher-awarded score. The moderators can be a panel of teachers, external experts, external examination board members, or board representatives.

The inspection can be by post. Moderators can visit a school, or converge at a centre. The procedure is known variously as group, consensus, external, or cross moderation, depending on the mode employed. It is essentially a quality control approach. However, if both the process and the products of assessment are discussed by groups of moderators and teachers with the view to arriving at shared understanding of the criteria or standards in operation, the procedure can be both quality control and quality assurance. When used in a quality assurance form, the process can enhance the professional development of teachers in student assessment.

There are, however, some difficulties in using the inspection approach to moderate scores. The problems revolve around (1) agreement on standards and stating them explicitly; (2) differences in modes of assessment, marking schemes, and relative weightings given to the components; and (3) vulnerability to a number of reliability and validity problems. There is also the practical problem of assembling everyone for a moderation session if the group moderation procedure is adopted. In addition, Burton and Linn (1993) explain that in the use of group moderation, the comparability of scores assigned depends on the development of a consensus among the professionals. In the case of individual moderation, the specificity of the criteria used by moderators may vary from moderator to moderator. Burton and Linn (1993) found that in some cases "the decision as to whether the teachers' grades were out of line was left largely to the moderator without very specific criteria delineated" (p.9). Furthermore, the procedure also calls for a definition of tolerance limits so that if scores fall outside of these tolerance limits they are adjusted. An important characteristic of the use of tolerance limits is that they tend to differentially penalize students from different schools whose grades fall just and further outside the limits. Thus, tolerance limits are also known to suffer from validity problems. It should also be mentioned that, group moderation can be time consuming and costly. The procedure requires substantial investment of time and resources. Most developing countries such as Ghana, Nigeria, and Sierra Leone, to mention a few, lack some of the required resources - financial and human - at present. This is the reason why the WAEC was dissuaded from its use in the wake of the implementation of the practice in the West African sub-region (Ademola, 1992) but rather

opted for statistical moderation. The procedure, however, is becoming more and more popular in some countries.

Statistical Moderation

Statistical moderation, usually, is a procedure that uses information from an external examination to adjust teacher-based assessment scores (Burton and Linn, 1993; Cohen and Deale, 1977; Harlen, 1994; Linn, 1993; Smith, 1978). This is usually a linear transformation of the teacher-based assessment scores within a school or a centre to follow the distribution of the external examination scores. The rationale behind statistical moderation as Burton and Linn (1993) put it is that

> The teacher-assessed components of the examination are likely to be more valid in terms of ranking the students, while the external exam is more suitable for establishing the relative standard of work across schools (p.18).

The two most common procedures used in statistical moderation are (1) scaling and (2) mapping. In scaling, "the marks from the internal assessment for each assessment are ... adjusted to give the same mean and standard deviation as the distribution of marks for the moderating instrument of the candidates in that group" (Smith, 1978, p. 23). In mapping,

> The results on the moderating instrument of all candidates from each centre or assessment group are ranked. The candidates are also ranked in the order determined by the internal assessment. The top candidate on the internal assessment is then given a mark equivalent to the top mark obtained in the group on the moderating instrument, the next highest moderating test mark is given to the candidate ranked second by the centre, and so on, down the rank order for the internal assessment (Smith, 1978, p.24).

It is interesting to note that in both approaches, the teacher's rank ordering of the students is unchanged. This implies an acceptance of the rank ordering of the teachers within schools. However, different teachers' marks may change the relative standing of students in the total distribution of scores pooled across schools.

Statistical moderation is based on criteria that supposedly apply to teacher-based assessments (internal assessments) and external examinations. First, is the criterion that the internal assessments must be conducted over a period of time and that essentially the same knowledge, skills and abilities are assessed by both the internal and external assessments (Cohen and Deale, 1977; Burton and Linn, 1993; Smith, 1978). This criterion calls for the condition that the internal and external scores should be correlated. The Department of Education and Science, Welsh Office's document on appropriateness of moderation methodology for General Certificate of Secondary Education (1985) states:

> If statistical moderation against an externally assessed component is to be used, it is essential that there is a satisfactory level of correlation between the internally assessed component being moderated and the externally assessed component used to moderate it (p.23).

However, the issue of a satisfactory level of correlation between teacher assessment and external examination scores is a tricky and contentious one. Too little overlap (correlation) will render the moderating examination unsuitable and too much overlap (multicollinearity) will cast doubt on the advisability of having both components as part of the same examination process. It is recommended that correlation coefficients below 0.50-0.60 are possibly too low (Smith, 1978; Cohen and Deale, 1977).

Second, the average grade of candidates from a particular school should be at the same levels, within statistical limits for both the internal and external assessments, and if they are not, adjustments must be made to the internal assessments to bring the average score within tolerance limits. This criterion assumes that there should be no significant differences in attainment of students on both internal and external components in any particular school (Burton and Linn, 1993).

A third criterion is that the external assessment should be reliable and be capable of being marked with a high degree of consistency.

Practically, the criteria are difficult to meet fully. Smith (1978), therefore, argued that because no examination can ever fully satisfy the above criteria, it is "probably unwise to place all one's faith in the moderation instrument and to adjust candidates' internally assessed marks in strict accordance with performance on it as it would be to accept the internally assessed marks without applying any kind of moderating technique" (p. 26). As a solution, he advocates a midway position in which action is taken only if scores on the internal and external examination are sufficiently different. That is, only if the internal assessment scores fall outside established tolerance limits. As indicated earlier in this paper, the use of tolerance limits has its problems. For instance, how far should a student's score be from the tolerance limit before it is subjected to moderation? Should different moderation procedures be applied to scores based on how far apart the score is from the tolerance limit?

The establishment of the tolerance limits is also affected by a number of factors: (1) the extent to which the external examination satisfies the above criteria, (2) the standard deviation of marks in the two assessments, and (3) the number of candidates taking each assessment (Burton and Linn, 1993; Smith, 1978).

From the foregoing, it is abundantly clear that the use of statistical moderation is fraught with unresolved issues. The issues include: What degree of relationship between the external examination scores and teacher assessment scores will result in good moderation of teacher assessment scores? What tolerance limit is appropriate for moderating teacher assessment scores? Do we moderate all scores from all schools or only some of them? Within schools, should only some subject area scores be moderated? The issues are further complicated by its use in a situation whereby the teacher assessments are supposedly criterion-referenced while the external assessment is norm-referenced.

The WAEC uses both scaling and the mapping methods in moderating the teacher-based assessment scores at the basic and secondary school levels (Ademola, 1992; Mansaray, 1988). The mapping procedure is used when the number of candidates in a school is small (Wuddah, 1996). In Ghana, however, only the scaling procedure is used to moderate the teacher-based continuous

assessment scores (personal communication with D. Odukoya, March 23, 1998).

Even though statistical moderation is used in some countries including New Zealand, Australia, Ghana, and Nigeria, its use has been discontinued in United Kingdom because of the difficulties associated with meeting the underlying assumption and criteria (Gipps, 1994; Harlen, 1994; Satterly, 1994).

Moderation by Monitoring

Another way of moderating teacher assessment scores is monitoring. This is a quality assurance procedure, which involves ensuring that assessments in the schools meet the expected national standards. In the literature, this type of moderation has been discussed under different headings such as national monitoring, visitation moderation, item banks, and monitoring procedure (Broadfoot, 1994; Harlen, 1994; New Zealand Qualifications Authority, 1992; Nuttall and Thomas, 1993). There are many shades in practice. Monitoring can take one or more of the following forms: (1) using items from question banks centrally generated for use by teachers (used to calibrate teachers' assessments), (2) using common assessment tasks, (3) auditing teacher assessments from time to time, (4) cross-moderating assessment tasks before administration, (5) having only accredited assessors use assessment instruments, and (6) using exemplar items.

Discussion

The main issues concerning combining teacher-based assessment scores with external examinations presented so far, namely, the quality of teacher assessments and external examinations and moderating teacher assessment scores are interrelated. Dealing with any one issue without considering the others may lead to some problems later. Thus a systemic approach needs to be adopted in the implementation of the practice. The relatively low reliability of teacher assessments depicted above behoves stakeholders in policy formulation regarding the combination to take a hard look at improving the competency of teachers in assessing students. In practical terms, it is possible to train teachers at the basic and secondary school levels to acquire fundamental skills in assessing students through pre-service and in-service courses. It may be unreasonable to require all teachers to acquire the very high

level skills in assessment required for educational assessment experts even though this may be desirable. It is true that the respective Ministries of Education are taking steps in improving teacher competency in assessing their students. What is needed now is to enhance, strengthen and sustain such efforts to ensure the competency of all teachers.

From the foregoing discussion of moderation of teacher assessment scores, it is obvious that quite a number of unresolved issues are associated with statistical moderation. One major issue is the relationship between the components. A study conducted by Adeyegbe (1993) on the relationship between continuous assessment and external examination scores in Nigeria, which can be meaningfully used as a proxy for Ghana, yielded correlation coefficients ranging from 0.24 to 0.86 for different schools and subjects. In English Language, for instance, the correlations ranged between 0.35 and 0.77 with the majority of the schools sampled (60%) having correlation coefficients below 0.50. The majority showed no statistically significant relationship between the pairs of scores. In mathematics, the correlation coefficients ranged between 0.44 and 0.86. All the coefficients showed a statistically significant relationship between the scores. Adeyegbe concluded that "generally speaking, there was not much relationship between CASS (continuous assessments) and TASS (external assessment) score" (p.179). Similar observations regarding the relationship between teacher assessment and external examination scores were made in Ghana (personal communication with K.E. Arthur of the Test Development and Research Unit of WAEC on January 2, 1997). In a study conducted by Amedahe (1998) involving 2,378 students from 11 randomly selected public senior secondary schools, the correlation coefficients between the teacher assessment scores and external examination scores in mathematics ranged between 0.42 and 0.86. In cases of high correlation coefficients (e.g. correlation coefficients above 0.55) between teacher assessment scores and external examination scores, the existence of multicolinearity may render the practice superfluous.

This and other problems led to its discontinued use in the United Kingdom and some other countries. It should be noted that moderating teacher assessment scores is not a panacea to any weaknesses that may

be inherent in the assessments. None of the procedures used to moderate teacher assessment scores is foolproof. Each may have its advantages and disadvantages as indicated in this paper.

The issues presented indicate, however, that there is the need to rather emphasize quality assurance procedures, such as monitoring and inspection of teacher assessments, rather than attempt to control the quality using statistical moderation, even though the procedures may be more difficult to practise. For example, when monitoring results in well-crafted and graded teacher assessments high-level quality scores may be obtained. Since the main weakness of teacher assessments is their lack of comparability from school to school and from teacher to teacher, the implementation of the assessment framework requires teachers to similarly assess and grade their students at specific grade levels. To be able to achieve the relatively high degree of comparability in assessment needed, it is important that the respective Ministries of Education provide the schools with detailed information on how the practice should be implemented. The present guidelines are not detailed enough. Item banks should be developed and used to ensure comparability of scores.

The author found out in a study in 1998 that at some senior secondary schools in Ghana, teachers at different schools computed students' term and year assessment scores differently. While some teachers computed an end of year score for a student in a subject based on number of tests, quizzes, class assignments, and homework, others computed the score based on end of term examinations alone. Even within schools there was inconsistency in the way the end of year scores were arrived at. It is also a fact that due to pressure of work, some teachers do not give students the recommended number of assessments. Instead, such teachers use their own judgment to assign scores to students. Such scores may or may not reflect the student's achievement in the subject area. Adequate supervision and monitoring should be put in place to ensure that teacher assessments meet the expectations. Since there is always room to improve in any human endeavour, the WAEC must also continue to search for ways to improve the quality of its examinations.

Conclusion

In conclusion, the practice of combining continuous assessment

scores with external examination scores for certifying students at the basic and secondary schools appear to be appealing. However, like any human endeavour, the practice is fraught with some issues such as quality of the assessments and moderation. These must be reviewed to ensure the validity and reliability of the scores. There is a need to take a hard look at the current mode of moderation of teacher assessment scores using a statistical procedure since the procedure has been discontinued in some countries and has been replaced by monitoring and inspection because of difficulties associated with its use.

References

Ademola, A. (1992, November). *The challenges of combining internal and external assessment in a certificate examination: The West African Examinations Council experience.* A paper presented at the Annual Meeting of American Evaluation Association, Seattle, WA.

Adeyebge, S.O. (1993). Correlation between continuous and terminal assessments at the senior school certificate examination (SSCE) level: The reality of a novel approach in the implementation of national policy on education. JORIC (7), 163-181.

AFT, NCME, & NEA. (1990). *Standards for teacher competence in educational assessment of students.* Washington DC: Author.

Amedahe, F.K. (1998). Models of combining continuous assessment scores with external examination scores for selection and certification. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh.

Amedahe, F. K. (1989). Testing practices in secondary schools in the Central Region of Ghana. Unpublished master's thesis, University of Cape Coast, Cape Coast, Ghana.

Angoff, W. H. (1971). Scales, Norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp.508-600). Washington, DC: American Council on Education.

Awuwoloye, E. O. (1988, September). *Problems of combining internal and external assessment.* A paper delivered at the International Conference on Education Policies and Systems, Ibadan, Nigeria.

Bardell, G.S., Forrest, G.M. & Shoesmith, D.J. (1978). Comparability in GCE: A review of the Boards' studies. 1964-1977. Manchester. JMB on behalf of the GEC Examination Boards.

Broadfoot, P. (1994). Approaches to quality assurance and control in six countries. In W. Harlen (Ed.). *Enhancing quality in assessment,* (pp.26-52). London: British Educational Research Association Policy Task Group on Assessment, Paul Chapman.

Burton, E. & Linn, R.L. (1993). Report on linking study-comparability across assessments: Lessons from the use of moderation procedures in England. Project 2.4: Quantitative models to monitor status and progress of learning and performance. ERIC No. ED364578.

Carter, K. (1984). Do teachers understand the principles for writing tests? *Journal of Teacher Education.* 35(6), 57-60.

Cohen, L. & Deale, R.N. (1977). Assessment by teachers in examinations at 16+. *Schools Council Examination Bulletin.* 37. London: Evans/ Methuen Educational.

Department of Education and Science. (1985). *General Certificate of Secondary Education: The National Criteria.* London: England, Department of Education and Science/Welsh Office.

Feldt, L.S. (1997). Can validity rise when reliability declines. *Applied Measurement in Education,* 10(4), pp. 377-387.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W.E. Hathaway (Ed.), *Testing in the schools: New directions for testing and measurement.* (pp.19-38). San Francisco: Jossey-Bass.

Frisbie, D.A. (1988). Reliability of scores from teacher-made tests: An NCME instructional module. *Educational Measurement: Issues and Practice*, 7(1), 25-33.

Gipps, C. (1994). Quality in teacher assessment. In W. Harlen (Ed.) *Enhancing quality in assessment,* (pp.7-86). London: British Educational Research Association Policy Task Group on Assessment, Paul Chapman.

Gullickson A.R. & Ellwein, M.C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. *Educational Measurement: Issues and Practice*, 4(1), 15-18.

Harlen, W. (1994), (Ed.), *Enhancing quality in assessment* (pp.11-25). London: British Educational Research Association Policy Task Group on Assessment, Paul Chapman.

Harris, A. (1986). *School-based assessment in GCE and CSE Boards: A report on policy and practice.* London: Secondary Examinations Council.

Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada.* Edmonton, Alberta: Author.

Kolen, M.J. & Brennan, R.L. (1995). *Test equating methods and practices.* New York: Springer-Verlag.

Linn, R.L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6 (1), 83-102.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum.

Mansary, M.L. (1988, September). *Statistical moderation under the Senior School Certificate Examination (SSCE) System revisited.* A paper delivered at International Conference on Education Policies and Systems, Ibadan, Nigeria.

Ministry of Education (1988). *Report of the sub-committee on assessment at basic education level.* Accra. Ghana: Author.

New Zealand Qualifications Authority (1992). *Designing a moderation system: Developing qualifications framework for New Zealand.* Wellington: Author.

Nuttall, D.L. & Thomas, S. (1993). *Monitoring procedures based on Centre performance variables.* Research and Development Series No. 11. Sheffield, England: Training, Enterprise and Education Directorate.

Opong, S. (1992). *Implementation of the Junior Secondary school program.* A paper presented at the Research Seminar on Primary Education in Ghana, University of Cape Coast, Cape Coast, Ghana.

Petersen, N.S., Kolen, M.J. & Hoover, H.D. (1998). Scaling, Norming, and Equating. In R.L. Linn (Ed.) *Educational measurement* (3rd ed.) (pp. 221-262). Phoenix: National Council on Measurement in Education and American Council on Education, Oryx Press.

Smith, G.A. (1978). *JMB experience of the moderation of internal assessments (occasional paper 38).* Manchester, England: Joint Matriculation Board.

Stiggins, R.J. (1994). *Student-centered classroom assessment.* New York, NY: Macmillan.

Stiggins, R.J. & Conklin, N.F. (1992). *In teachers hands: Investigating the practices of classroom assessment.* Albany, NY: State University of New York Press.

Wuddah, A.A. (1996, July). *Incorporating teacher assessment in the final grade.* A paper presented at the WAEC Monthly Seminar, Accra, Ghana.