Self and Peer Assessment in School and University: Reliability, Validity and Utility

Keith Topping
Department of Psychology, University of Dundee, Scotland

1. INTRODUCTION

Self-assessment and peer assessment might appear to be relatively new forms of assessment, but in fact, they have been deployed in some areas of education for many years. For example, George Jardine, professor at the University of Glasgow from 1774 to 1826, described a pedagogical plan including methods, rules and advantages of peer assessment of writing (Gaillet, 1992). By 1999, Hounsell and McCulloch noted that over a quarter of assessment initiatives in a survey of higher education (HE) institutions involved self and/or peer assessment. Substantial reviews of the research literature on self and peer assessment have also appeared (Boud, 1995; Boud & Falchikov, 1989; Brown & Dove, 1991; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000; Topping, 1998).

Why should teachers, teacher educators and education researchers be interested in these developments? Can they enhance quality and/or reduce costs? Do they work? Under what conditions? This chapter explores the conceptual underpinnings and empirical evidence for the reliability, validity, effects, utility, and generalizability of self and peer assessment in schools and higher education, and by implication in the workplace and lifelong learning.

All forms of assessment should be fit for their purpose, and the purpose of any assessment is a key element in determining its validity and/or reliability. The nature and purposes of assessments influence many facets of

M. Segers et al. (eds.),
Optimising New Modes of Assessment: In Search of Qualities and Standards, 55–87.

student performance, including anxiety (Wolfe & Smith, 1995), goal orientation (Dweck, 1986), and perceived controllability (Rocklin, O'Donnell, & Holst, 1995). Of course, different stakeholders might have different purposes.

Many teachers successfully involve students collaboratively in learning and thereby relinquish some control of classroom content and management. However, some teachers might be anxious about going so far as to include self or peer assessments as part of summative assessment, where consequences follow from terminal judgements of accomplishments. By contrast, formative or heuristic assessment is intended to help students plan their own learning, identify their own strengths and weaknesses, target areas for remedial action and develop meta-cognitive and other personal and professional transferable skills (Boud, 1990, 2000; Brown & Knight, 1994). Triangulating formative feedback through the inclusion of self and peer assessment might seem to incur fewer potential threats to quality.

Reviews have confirmed the utility of formative assessment (e.g., Crooks, 1988), emphasising the importance of quality as well as quantity of feedback. Black & Wiliam (1998) concluded that assessment which precisely indicated strengths and weaknesses and provided frequent constructive individualised feedback led to significant learning gains, as compared to traditional summative assessment. The active engagement of learners in the assessment process was seen as critical, and self-assessment an essential tool in self-improvement. Affective aspects, such as the motivation to respond to feedback and the belief that it made a difference, were also important.

However, the new rhetoric on assessment might not be matched by professional practice. For example, MacLellan (2001) found that while university staff declared a commitment to the formative purposes of assessment and maintained that the full range of learning was frequently assessed, they actually engaged in practices which militated against formative and authentic assessment being fully realised.

Explorations of self and peer assessment might be driven by a need to improve quality or a need to reduce costs. These two purposes are often intertwined, since a professional assessor confronted with twice as many products to assess in the same time is likely to allocate less time to each unit of assessment, with consequent implications for the reliability and validity of the professional assessment. A peer assessor with less skill at assessment but more time in which to do it might produce an equally reliable and valid assessment. Peer feedback might be available in greater volume and with greater immediacy than teacher feedback, which might compensate for any quality disadvantage.

Beyond education, self and peer assessment (or self-improvement through self-evaluation prior to peer evaluation) are increasingly found in workplace settings (e.g. Bernadin & Beatty, 1984; Farh, Cannella, & Bedeian 1991; Fedor & Bettenhausen, 1989; Joines & Sommerich, 2001), sometimes in the guise of "Total Quality Management" or "Best Value" exercises (e.g., Kaye & Dyason, 1999). The development of such skills in school and HE should thus be transferable. University academics have long been accustomed to peer assessment of submissions to journals and conferences, the reliability and validity of which has been the subject of empirical investigation (and some concern) for many years (e.g., Cicchetti, 1991). Teachers, doctors and other professionals are often assessed by peers in vivo during practice. All of us may expect to be peer assessor and peer assessee at different times and in different contexts - or as Cicchetti (1982) more colourfully phrased it in a paper on peer review: "we have met the enemy and he is us" (p. 205).

Additionally, peer assessment in particular is connected with other forms of peer assisted learning in schools and HE. Recent research has considerably clarified the many possible varieties of peer assisted learning, their relative effectiveness in a multiplicity of contexts, and the organisational parameters crucial for effectiveness (Boud, Cohen, & Sampson, 2001; Falchikov, 2001; Topping, 1996a,b; 2001a,b; Topping & Ehly, 1998).

In this chapter, self-assessment is considered first, then peer assessment. For each practice, a definition and typology of the practice is offered, followed by a brief discussion of its theoretical underpinnings. The "accuracy", reliability and validity of the practice in schools and higher education is then considered. The research findings of the effects of the practice are then reviewed in separate sections focused on schools and higher education respectively. The research literature was searched online and manually and all relevant items included in the database for this systematic review, which consequently should have no particular bias (although space constraints do not permit mention of every relevant study by name). A summary and conclusions section for each practice relates and synthesises the findings. Finally, studies directly comparing self and peer assessment are considered, followed by an overall summary and conclusion encompassing and comparing and contrasting both practices. Evidence-based guidelines for quality implementation of self and peer assessment are then given.

2. SELF ASSESSMENT

2.1 Self Assessment - Definition, Typology and Purposes

Assessment is the determination of the amount, level, value or worth of something. Self-assessment is an arrangement for learners and/or workers to consider and specify the level, value or quality of their own products or performances.

In self-assessment, the intention is usually to engage learners as active participants in their own learning and foster learner reflection on their own learning processes, styles and outcomes. Consequently, self-assessment is often seen as a continuous longitudinal process, which activates and integrates the learner's prior knowledge and reveals developmental pathways in learning. In the longer term, it might impact self-management of learning - facilitating continuous adaptation, modification and tuning of learning by the learner, rather than waiting for others to intervene. There is evidence that graduates in employment regard the ability to evaluate one's own work as a crucial transferable skill (e.g., Midgley & Petty, 1983).

There is a large commercial market in the publication of self-test materials or self-administered revision quizzes. These are often essentially rehearsal for external summative assessment, are not used under controlled or supervised conditions, do not appear to have been rigorously evaluated, seem likely to promote superficial, mechanistic and instrumental learning, and are not our focus here. However, computerised curriculum-based self assessment test programmes which give continuous rich formative feedback to learners (often termed "Learning Information Systems") have been found effective in raising student achievement in schools (e.g., Topping, 1999; Topping & Sanders, 2000).

Self-assessment operates in many different curriculum areas or subjects. The products, outputs or performances assessed can vary - writing, portfolios, oral and/or audio-visual presentations, test performances, other skilled behaviours, or combinations of these. Where skilled behaviours in professional practice are self-assessed, this might occur via retrospective recollection or by post hoc analysis of video recordings. The self-assessment can be summative (judging a final product or performance to be correct/incorrect or pass/fail, or assigning some quantitative mark or grade) and/or (more usually) formative (involving detailed qualitative assessment of better and worse aspects, with implications for making specific onward improvements). It may be absolute (referred to external objective benchmark criteria) or relative (referring to position in relation to the products or performances of the current peer group). Boud (1989) explores the issue of

whether self-assessment should form part of official student gradings, controversial if the practice is assumed to be of uncertain reliability and validity, and raising concerns about issues of power and control.

2.2 Self Assessment - Theoretical Underpinnings

What does self assessment require from students in terms of cognitive, meta-cognitive and social-affective demands? Through what processes might these benefit students? Under what conditions these processes might be optimised?

Self-assessment shares some of the characteristics of peer assessment, the theoretical underpinnings of which are discussed in detail later. Any form of assessment is a cognitively complex undertaking, requiring understanding of the goals of the task(s), the criteria for success and the ability to make judgements about the relationship of the product or performance to these. The process of self-assessment incurs extra time on task and practice. It requires intelligent self-questioning - itself cognitively demanding - and is an alternative structure for engagement with learning which seems likely to promote post hoc reflection. It emphasises learner ownership and management of the learning process, and seems likely to heighten the learner's sense of personal accountability and responsibility, as well as motivation and self-efficacy (Rogers, 1983; Schunk, 1996). All of these features are likely to enhance meta-cognition. At first sight self-assessment might seem a lonelier activity than peer assessment, but it can lead to interaction, such as when discussing assessment criteria or when the learner is called upon to justify their self-assessment to a peer or professional tutor. Such onward discussions involve constructing new schemata, moderation, norm-referencing, negotiation and other social and cognitive demands related to the mindful reception and assimilation of feedback.

2.3 Self Assessment - Reliability and Validity

This section considers the degree of correspondence between student self-assessments and the assessments made of student work by external "experts" such as professional teachers. This might be termed "accuracy" of self-assessment, if one assumes that expert assessments are themselves highly reliable and valid. As this is a doubtful assumption in some contexts (see below), it is debatable whether studies of such correspondence should be considered to be studies of reliability or validity or both or neither. This confusion is reflected in the very various vocabulary used in the literature.

There is evidence that the assessment of student products by professionals is very variable (Heywood, 1988; Newstead & Dennis, 1994; Newstead, 1996; Rowntree, 1977). Inter-rater reliabilities have been shown to vary from 0.40 to 0.63 (fourth- and eighth-grade writing portfolios) (Koretz, Stecher, Klein, & Mc Caffrey, 1994), through 0.58 to 0.87 (middle and high school writing portfolios) (LeMahieu, Gitomer, & Eresh, 1995) and 0.68 to 0.73 (elementary school writing portfolios) (Supovitz, MacGowan, & Slattery, 1997), to 0.76 to 0.94 (elementary school writing portfolios) (Herman, Gearhart, & Baker, 1993), varying with the dimensions assessed and grade level. This context should condition expectations for the "reliability" and "validity" of assessments by learners, in which the developmental process is arguably more important than "accuracy". However, Longhurst and Norton (1997) showed that tutor grades for an essay correlated quite highly (0.69 - 0.88) with deep processing criteria, while the correlation between student and tutor grades was lower (0.43).

For schoolchildren, Barnett and Hixon (1997) found age and subject differences in the reliability of self-assessment in school students. Fourth graders made relatively accurate predictions in each of three subject areas. Second graders were similar except for poor predictions in mathematics. Sixth graders made good predictions in mathematics and social studies, but not in spelling. Blatchford (1997) found race and gender differences in the reliability of self assessment in school pupils aged 7-16 years. White pupils were less positive about their own attainments and about themselves at school. While black girls showed confidence in their attainments, and had the highest attainments in reading and the study of English, white girls tended to underestimate themselves and have little confidence.

In higher education, Falchikov and Boud (1989) reported a meta-analysis of self-assessment studies which compared teacher and student marks. The degree of correspondence varied widely in different studies, from a low correlation coefficient of -0.05 to a high of 0.82, with a mean of 0.39. Some studies gave inter-assessor agreement as a percentage, and this varied from 33% to 99%, with a mean of 64%. Correspondence varied with: design and implementation quality of the study (better studies showing higher correspondence), level of the course (more advanced learners showing higher correspondence), area of study (science subjects showing higher correspondence than social science), and nature of product or performance (academic products showing higher correspondence than professional practice). Self-assessments focusing on effort rather than achievement were particularly unreliable. Overall, self-assessed grades tended to be higher than staff grades. However, more advanced students tended to under-estimate themselves.

Boud and Falchikov (1989) conducted a critical analysis of the literature on student self-assessment in HE published between 1932 and 1988. The methodological quality of studies was generally poor, although later studies tended to be better. Some studies made no mention of any explicit criteria. Where there were criteria, very many different scales were used. Some studies included ratings of student effort (of very doubtful reliability). Self-assessment sometimes appeared to be construed as the learner's guess at the professional staff assessment, rather than a rationally based independent estimate. The context for the learning to be assessed was often insufficiently described. Reports of replications were rare.

There was a tendency for earlier studies to report self-assessor over-rating and later studies under-rating. Overall, more able students tended to under-rate themselves, and weaker students to over-rate themselves by a larger amount. An interesting exception (see Gaier, 1961), found that high and low ability students produced more accurate self-assessments than middle ranking students. Boud and Falchikov (1989) found that students in the later years of courses and graduates tended to generate self-assessments more akin to staff assessments than those of students early in courses. However, those longitudinal studies which allowed scrutiny of the impact of practice in self-assessment over time showed mixed results, four studies showing improvement, three studies no improvement. Studies of any gender differences were inconclusive.

More recently, Zoller and Ben-Chaim (1997) found that students over-estimated not only their abilities in the subject at hand, but also their abilities in self-assessment, as compared to tutor assessments. A review of self-assessment in medical education concluded that despite the accepted theoretical value of self-assessment, the reliability of the procedure was poor (Ward, Gruppen, & Regehr, 2002). However, several later studies have shown that the ability of students to assess themselves improves in the light of feedback or with time (Birenbaum & Dochy, 1996; Griffee, 1995). Frye, Richards, Bradley and Philp (1992) found individual students had a tendency towards over- or under-estimation in prediction of examination performance that was relatively consistent, but evolved over time with experience, maturity and self-assessment practice towards decreased overestimation and increased underestimation. Ross (1998) summarised research on self assessment, meta-analysing 60 correlations reported in the second-language testing literature. Self-assessments and teacher assessments of recently instructed ESL learners' functional English skills revealed differential validities for self-assessment and teacher assessment depending on the extent of learners' experience with the self-assessed skill.

2.4 Self Assessment in Higher Education: Effects

In considering the effects of self-assessment, the question arises of "what is a good result?" A finding that learners undertaking self-assessment have better outcomes than learners who do not, other things being equal, is clearly a "good result". A finding that learners undertaking self-assessment instead of professional assessment have outcomes as good as (if not significantly better than) learners receiving professional assessment is also arguably a "good result". However, a finding that learners undertaking self-assessment in addition to professional assessment have outcomes only as good as (and not significantly better than) learners receiving only professional assessment is not a "good result".

There are relatively few empirical studies of the effects of self-assessment. Davis and Rand (1980) compared the performance of an instructor-graded and a self-graded class. Although the self-graded class over-estimated, their overall performance was the same as the instructor-graded class. This suggests that the formative effects of self-assessment are no less than those of instructor grading, with much less effort on the part of the instructor. Sobral (1997) evaluated self-assessment of elective self-directed learning tasks, finding increased levels of self-efficacy and significant relationships to measures of deep approaches to study. Academic achievement (Grade Point Average) was significantly higher for experimental students than for controls, although not all experimental students benefited.

Marienau (1999) found longitudinal perceptions among adult learners that the experience of self-assessment strengthened commitment to subsequent competent performance, enhanced higher order skills, and fostered self-direction, illustrating that effects might not necessarily be immediate. El-Koumy (2001) investigated the effects of self-assessment on the knowledge and academic thinking of 94 English as a Foreign Language (EFL) students. Students were randomly assigned to experimental and control groups. The self-assessment group was required to assess their own knowledge and thinking before and after each lecture, during a semester. Both groups were pre- and post-tested on knowledge and academic thinking. The experimental group scored higher on both, but differences did not reach statistical significance.

2.5 Self Assessment in Schools: Effects

Similar caveats about "what is a good result?" apply here. Towler and Broadfoot (1992) reviewed the use of self-assessment in the primary school. They argued that assessment should mainly be the responsibility of the

learner, and that this principle could be realistically applied in education from the early years, while emphasising the need for pupil training and a whole school approach to ensure quality and consistency. Self-assessment has indeed been successfully undertaken with some rather unlikely populations in schools, including students with learning disabilities (e.g., Lee, 1999; Lloyd, 1982; Miller, 1988) and pre-school and kindergarten children (e.g., Boersma, 1995; Mills, 1994).

Lloyd (1982) compared the effects of self assessment and self-recording as interventions for increasing the on-task behaviour and academic productivity of elementary school learning disabled students aged 9-10 years. For this population, self-recording appeared a more effective procedure than self-assessment. Miller (1988) noted that learning handicapped students tend to be passive learners. For them, self assessment included "sizing up" the task before beginning, gauging their own skill level and likelihood of success before beginning, continuous self-monitoring and assessment during task performance, and consideration of the quality of the final product or performance. Self-assessment effectiveness was seen as likely to vary according to three sets of parameters: person variables (such as age, sex, developmental skills, self-esteem), task variables (such as meaningfulness, task format, level of complexity), and strategy variables (specific strategy knowledge, relational strategy knowledge, and meta-memory).

Even with pre-school children, portfolios can be used to develop the child's own self-assessment skills and give a focus to discussions between the child and salient adults. In Mills' (1994) study, portfolios were organised around four areas of development: physical, social and emotional, emergent literacy, and logico-mathematical. For each area there was a checklist, and evidence was included to back up the checklist. At points during the school year, a professional met with each parent to discuss the portfolio, results, and goals for the child. The portfolio was subsequently made available to the child's kindergarten teacher.

Boersma (1995) described curricular modifications designed to increase students' ability to self-assess and set goals in grades K-5. Problems with self-evaluation and goal setting were documented through parent, teacher, and student surveys. Interventions included the development of a portfolio system of assessment and the implementation of reflective logs and response journals. These were successful in improving student self-evaluation and goal setting across the grades, but improvement was more marked for the older students.

Rudd and Gunstone (1993) studied the development of self-assessment skills in science and technology in third grade children. Self-assessment was scaffolded through questionnaires, concept maps and graphs created by

students. Specific self-assessment concepts and techniques were introduced to the students during each term, over one academic year. Student awareness and use of skills in these classes were substantially enhanced. The teacher's role changed from controller to delegator as students became more proficient at self-assessment.

There is some evidence that engagement in self-assessment has positive effects on achievement in schools. Sink, Barnett and Hixon (1991) found that planning and self-assessment predicted higher academic achievement in middle school students. Fontana and Fernandes (1994) tested the effects of the regular use of self-assessment techniques on mathematical performance with children in 25 primary school classes. Children (n=354) in these classes showed significant improvements in scores on a mathematics test, compared to a control group (n=313). In a replication, Fernandes and Fontana (1996) found children trained in self-assessment showed significantly less dependence upon external sources of control and upon luck as explanations for school academic events, when compared to a matched control group. In addition, the experimental children showed significant improvements in mathematics scores relative to the control group.

Ninness, Ninness, Sherman and Schotta (1998) and Ninness, Ellis and Ninness (1999) trained school students in self-assessment by computer-interactive tutorials. Students received computer-displayed accuracy feedback plus reinforcement for correct self-assessments of their math performance. After withdrawal of reinforcement, self-assessment alone was found motivational, facilitating high rates and long durations of math performance. McDonald (2002) gave experimental high school students extensive training in self-assessment and using a post-test only design compared their subsequent public examination performance to that of controls, finding the self-assessment group superior.

Additionally, self-assessment in schools is not confined to academic progress. Wassef, Mason, Collins, O'Boyle and Ingham (1996) evaluated a self-assessment questionnaire for high school students on emotional distress and behavioural problems, and found it reliable in relation to staff perceptions.

2.6 Summary and Conclusions on Self Assessment

Self-assessment is increasingly widely operated in schools and HE, including with very young children and those with special educational needs or learning disabilities. It is widely assumed to enhance meta-cognition and self directed learning, but this is unlikely to be automatic. The solid evidence for this is small, although encouraging. It suggests self-assessment can result in gains in learner management of learning, self-efficacy, deep rather than

superficial learning, and on traditional summative tests. Effects have been found to be at least as good as those from instructor assessment, and often better. However, effects might not be immediate and might be cumulative.

The reliability and validity of instructor assessment is not high, but that of self-assessment tends to be a little lower and more variable, with a tendency to over-estimation. The reliability and validity of self-assessment tends to be higher in relation to the ability of the learner, the amount of scaffolding, practice and feedback and the degree of advancement in the course, rather than chronological age. Other variables affecting reliability and validity include: the nature of the subject area, the nature of the product or performance assessed, the nature and clarity of the assessment criteria, the nature of assessment instrumentation, and cultural and gender differences.

In all sectors, much further development is needed, with improved implementation and evaluation quality and fuller and more detailed reporting of studies. Exploration of the effects of self-assessment is particularly needed.

3. PEER ASSESSMENT

3.1 Peer Assessment: Definition, Typology & Purposes

Assessment is the determination of the amount, level, value or worth of something. Peer assessment is an arrangement for learners and/or workers to consider and specify the level, value or quality of a product or performance of other equal-status learners and/or workers.

Peer assessment activities can vary in a number of ways, operating in different curriculum areas or subjects. The product or output to be assessed can vary - writing, portfolios, oral presentations, test performance, or other skilled behaviours. The peer assessment can be summative or formative. The participant constellation can vary - the assessors may be individuals or pairs or groups; the assessed may be individuals or pairs or groups. Directionality can vary - peer assessment can be one-way, reciprocal or mutual. Assessors and assessed may come from the same or different year of study, and be of the same or different ability. Place and time can vary - peer assessment can be formal and in class, or occur informally out of class. The objectives for the exercise may vary - the teacher may target cognitive or meta-cognitive gains, time saving, or other goals.

3.2 Peer Assessment - Theoretical Underpinnings

What does peer assessment require from students in terms of cognitive, meta-cognitive and social-affective demands? Through what processes might these benefit students? Under what conditions might these processes be optimised?

3.2.1 Feedback

The conditions under which feedback in learning is effective are complex (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler & Winne, 1995; Kulhavy & Stock, 1989). Feedback can reduce errors and have positive effects on learning when it is received thoughtfully and positively. It is also essential to the development and execution of self-regulatory skills (Bangert-Drowns, et al., 1991; Paris & Newman, 1990; Paris & Paris, 2001). Butler and Winne (1995) argue that feedback serves several functions: to confirm existing information, add new information, identify errors, correct errors, improve conditional application of information, and aid the wider restructuring of theoretical schemata. Students react differently to feedback from peers and from adults (Cole, 1991; Dweck & Bush, 1976; Henry, 1979). Gender differences in responsiveness to peer feedback have also been found (Dweck & Bush, 1976), but this interacts with age (Henry, 1979).

3.2.2 Cognitive Demands

Providing effective feedback or assessment is a cognitively complex task requiring understanding of the goals of the task and the criteria for success, and the ability to make judgements about the relationship of the product or performance to these. Webb (1989) and Webb and Farivar (1994) identified conditions for effective helping: relevance to the goals and beliefs of the learner, relevance to the particular misunderstandings of the learner, appropriate level of elaboration, timeliness, comprehension by the help-seeker, opportunity to act on help given, motivation to act, and constructive activity.

Cognitively, peer assessment might create effects by gains in a number of variables pertaining to cognitive challenge and development, for assessors, assessees, or both (Topping & Ehly, 1998, 2001). These could include levels of time on task, engagement, and practice, coupled with a greater sense of accountability and responsibility. Formative peer assessment is likely to involve intelligent questioning, coupled with increased self-disclosure and thereby assessment of understanding. Peer assessment could enable earlier error and misconception identification and analysis. This could lead to the

identification of knowledge gaps, and engineering their closure through explaining, simplification, clarification, summarising and cognitive restructuring. Feedback (corrective, confirmatory, or suggestive) could be more immediate, timely, and individualised. This might increase reflection and generalisation to new situations, promoting self-assessment and greater meta-cognitive self- awareness. Cognitive and meta-cognitive benefits might accrue before, during or after the peer assessment. Falchikov (1995, 2001) noted that "sleeper" (delayed) effects are possible.

3.2.3 Social Demands

Any group can suffer from negative social processes, such as social loafing, free rider effects, diffusion of responsibility, and interaction disabilities (Cohen, 1982; Salomon & Globerson, 1989). Social processes might influence and contaminate the reliability and validity of peer assessments (Byard, 1989; Falchikov, 1995; Pond, Ul-Haq, & Wade, 1995). Falchikov (2001) explores questions of role ambiguity, dissonance and conflict in relation to authority and status issues and attribution theory. Peer assessments might be partly determined by: friendship bonds, enmity or other power processes, group popularity levels of individuals, perception of criticism as socially uncomfortable or even socially rejecting and inviting reciprocation, or collusion leading to lack of differentiation. The social influences might be particularly strong with "high stakes" assessment, for which peer assessments might drift toward leniency (Farh, et al., 1991). Magin (2001 a) noted that concerns about peer assessment are often focused upon the potential for bias emanating from social considerations - so-called "reciprocity effects". However, in his own study he found such effects accounted for only 1 % of the variance. In any case, all these social factors require professional teacher scrutiny and monitoring. However, peer assessment demands social and communication skills, negotiation and diplomacy (Riley, 1995), and can develop teamwork skills. Learning how to give and accept criticism, justify one's own position and reject suggestions are all useful transferable social and assertion skills (Marcoulides & Simkin, 1991).

3.2.4 Affect

Both assessors and assessees might experience initial anxiety about the process. However, peer assessment involves students directly in learning, and might promote a sense of ownership, personal responsibility and motivation. Giving positive feedback first might reduce assessee anxiety and improve acceptance of negative feedback. Peer assessment might also

increase variety and interest, activity and inter-activity, identification and bonding, self-confidence, and empathy with others - for assessors, assessees, or both.

3.2.5 Systemic Benefits

Peer assessment offers triangulation and per se seems likely to improve the overall reliability and validity of assessment. It can also give students greater insight into institutional assessment processes (Fry, 1990), perhaps developing greater tolerance of inevitable difficulties of discrimination at the margin. It has been contended that peer assessment is not costly in teacher time. However, other authors (e.g., Falchikov, 2001) caution that there might be no saving of time in the short to medium term, since establishing good quality peer assessment requires time for organisation, training and monitoring. If the peer assessment is to be supplementary rather than substitutional, then no saving is possible, and extra costs or opportunity costs will be incurred. However, there might be meta-cognitive benefits for staff as well as students. Peer assessment can lead staff to scrutinise and clarify assessment objectives and purposes, criteria and marking scales.

3.3 Peer Assessment - Reliability and Validity

This section considers the degree of correspondence between student peer assessments and the assessments made of student work by external "experts" such as professional teachers. Caveats regarding the use of the terms "accuracy", "reliability" and "validity" are as for self-assessment. Many purported studies of "reliability" might be considered studies of "accuracy" or "validity", comparing peer assessments with assessments made by professionals, rather than with those of other peers, or the same peers over time.

Additionally, many studies compare marks, scores and grades awarded by peers and staff, rather than upon more open-ended formative feedback. This raises concerns about the uncertain psychometric properties of such scoring scales (such as sensitivity and scalar properties), alignment of the mode of assessment with teaching and learning outcomes (i.e. relevance of the assessment), and consequently validity in any wider sense. By contrast, the reliability and validity of detailed formative feedback was explored by Falchikov (1995) and Topping, Smith, Swanson, & Elliot (2000), for example.

Research findings on the reliability and validity of peer assessment mostly emanate from studies in HE. In a wide variety of subject areas and years of study, the products and performances assessed have included:

essays (Catterall, 1995; Haaga, 1993; Marcoulides & Simkin, 1991, 1995; Orpen, 1982; Pond, et al., 1995), hypermedia creations (Rushton, Ramsey, & Rada, 1993), oral presentations (Freeman, 1995; Hughes & Large, 1993a,b; Magin & Helmore, 2001), multiple choice test questions (Catterall, 1995), practical reports (Hughes, 1995), individual contributions to a group project (Mathews, 1994; Mockford, 1994) and professional skills (Korman & Stubblefield, 1971; Ramsey, Carline, Blank, & Wenrich, 1996). Methods for computerising peer assessment are now appearing (e.g., Davies, 2000).

Over 70% of the HE studies find "reliability" and "validity" adequate, while a minority find these variable (Falchikov & Goldfinch, 2001; Topping, 1998). MacKenzie (2000) reported satisfactory reliability for peer assessment of performance in viva examinations. Magin & Helmore (2001) found inter-rater reliability for tutors making parallel assessments of oral presentations higher than that for peer assessments, but the reliability for tutors was not high (0.40 to 0.53). Magin & Helmore (2001) concluded that the reliability of summative assessments of oral presentations could be improved by combining teacher marks with the averaged marks obtained from multiple peer ratings. A tendency for peer marks to bunch around the median is sometimes noted (e.g., Catterall, 1995; Taylor, 1995). Student acceptance (or belief in reliability) varies from high (Falchikov, 1995; Fry, 1990; Haaga, 1993) to low (Rushton, et al., 1993), quite independently of actual reliability.

Contradictory findings might be explained in part by differences in contexts, the level of the course, the product or performance being evaluated, the contingencies associated with those outcomes, clarity of judgement criteria, and the training and support provided. Reliability tends to be higher in advanced courses; lower for assessment of professional practice than for academic products. Discussion, negotiation and joint construction of assessment criteria with learners is likely to deepen understanding, give a greater sense of ownership, and increase reliability (see the review by Falchikov & Goldfinch, 2000 - although Orsmond, Merry and Reiling, 2000, found otherwise). Reliability for an aggregated global peer mark might be satisfactory, but not for separate detailed components (e.g., Lejk & Wyvill, 2001; Magin, 2001 b; Mockford, 1994). Peer assessments are generally less reliable when unsupported by training, checklists, exemplification, teacher assistance and monitoring (Lawrence, 1996; Pond, et al., 1995; Stefani, 1992, 1994). Segers and Dochy (2001) found peer marks correlated well with both tutor marks and final examination scores.

Findings from HE settings might not apply in other contexts. However, a number of other studies in the school setting have found encouraging consistency between peer and teacher assessments (Karegianes, Pascarella,

&, Pflaum, 1980; Lagana, 1972; MacArthur, Schwartz, & Graham, 1991; Pierson, 1967; Weeks & White, 1982).

3.4 Peer Assessment in Schools: Effects

Similar caveats about "what is a good result?" apply to peer assessment as to self-assessment. In schools, much peer assessment has focused on written products or multimedia work portfolios. A review has been provided by O'Donnell and Topping (1998).

3.4.1 Peer Assessment of Writing

Peer assessment of writing might involve giving general feedback, or going beyond that to very specific feedback about possible improvements. Peer assessment can focus on the whole written product, or components of the writing process, such as planning, drafting or editing. Studies in schools have shown less interest in reliability and validity than in HE, and more interest in effects on subsequent learner performance. Peer assessment of writing is also used with classes studying English as a Second or Additional Language (ESL, EAL) and foreign languages (Byrd, 1994; Samway, 1993).

Bouton and Tutty (1975) reported a study of the effects of peer assessment of writing with high school students. The experimental group did better than control group in a number of areas. Karegianes, et al. (1980) examined the effects of peer editing on the writing proficiency of 49 low-achieving tenth grade students. The peer edit group had significantly higher writing proficiency than students whose essays were edited by teachers. Weeks and White (1982) compared groups of grade 4 and 6 students in peer editing and teacher editing conditions. Differences were not significant, but the peer assessment group showed more improvement in mechanics and in the overall fluency of writing.

Raphael (1986) compared peer editing and teacher instruction with fifth and sixth grade students and their teachers, finding similar improvements in composition ability. Califano (1987) made a similar comparison in two fifth and two sixth grade classes, with similar results in writing ability and attitudes toward writing. Cover's (1987) study of peer editing with seventh graders found a statistically significant improvement in editing skills and attitudes toward writing. Wade (1988) combined peer feedback with peer tutoring for sixth-grade students. After training, the children could provide reliable and correct feedback, and results clearly demonstrated improvements in student writing.

Holley (1990) found peer editing of grammatical errors with grade 12 high school students in Alabama resulted in a reduction in such errors and

greater student interest and awareness. MacArthur and his colleagues (1991) used peer editing in grades 4-6 in special education classrooms, which proved more effective than only regular teacher instruction. Stoddard and MacArthur (1993) demonstrated the effectiveness of peer editing with seventh and eighth grade students with learning disabilities. The quality of writing increased substantially from pre-test to post-test, and the gains were maintained at follow-up and generalised to other written work.

3.4.2 Peer Response Groups

Peer response groups are a group medium for peer assessment and feedback, obviously involving different social demands to peer assessment between paired individuals. Gere and Abbot (1985) analysed the quality of talk in response groups. Students did stay on task and provide content-related feedback. Younger students spent more time on content than did older students, who attended more to the form of the writing. However, when Freedman (1992) analysed response groups in two ninth grade classrooms, she concluded that students suppressed negative assessments of their peers.

The effects of revision instruction and peer response groups on the writing of 93 sixth grade students were compared by Olson (1986, 1990). Students receiving instruction that included both teacher revision and peer assessment wrote rough and final drafts which were significantly superior to those of students who received teacher revision only, while peer assessment only students wrote final drafts significantly superior to revision instruction only students. Rijlaarsdam (1987) and Rijlaarsdam and Schoonen (1988) made similar comparisons with 561 ninth grade students in eight different schools. Teacher instruction and peer assessment proved equally effective.

Weaver (1995) surveyed over 500 instructors about peer response groups in writing. Regardless of the stage in the writing process (early vs. late), instructors generally found peer responses to be more effective than the teacher's. In contrast, students stated they found the teacher's responses to be more helpful in all stages of writing. Nevertheless, when students could seek peer responses at the Writing Centre but not in class, their writing improved.

3.4.3 Portfolio Peer Assessment

A portfolio is "a purposeful collection of student work that exhibits the student's efforts, progress, or achievement in one or more areas. The collection must include student participation in selecting contents, the criteria for judging merit, and evidence of the student's self reflection" (Paulson, Paulson, & Meyer, 1991, p. 60). Thus, a student must be able to

judge the quality of his or her own work and develop criteria for what should be included in order to develop an effective portfolio. However, there is as yet little adequate empirical literature on the effects of peer assessment of portfolios in schools.

3.4.4 Other Kinds of Peer Assessment in Schools

McCurdy and Shapiro (1992) deployed peers to undertake curriculum-based measurement in reading among 48 elementary students with learning disabilities, comparing with teacher and self-assessment. It was found that students in the self and peer conditions could collect reliable data on the number of correct words per minute. No significant differences were found between conditions. Salend, Whittaker and Reeder (1993) examined the efficacy of a consensus based group evaluation system with students with disabilities. The system involved: (a) dividing the groups into teams; (b) having each team agree on a common rating for the group's behaviour during a specified time period; (c) comparing each team's rating to the teacher's rating; and (d) delivering reinforcement to each team based on the group's behaviour and the team's accuracy in rating the group's behaviour. Results indicated that the system was an effective strategy for modifying behaviour. Ross (1995) had grade 7 students assess audio tape recordings of their own math co-operative learning groups at work. Increases in the frequency and quality of help seeking and help giving and improved students' attitudes about asking for help resulted.

3.5 Peer Assessment in Higher Education: Effects

Similar caveats about "what is a good result?" apply to peer assessment in HE as to self-assessment. In this section, studies of quantitative peer assessment are considered first, then other studies are grouped according to the product or performance assessed.

3.5.1 Peer Assessment through Tests, Marks or Grades

Hendrickson, Brady and Algozzine (1987) compared individually administered and peer mediated tests, finding scores significantly higher under the peer mediated condition. The latter was preferred by students, who found it less anxiety-provoking. Ney (1989) applied peer assessment to tests and mid-term and final exams. This resulted in improved mastery of the subject matter, and better classroom attendance. Stefani (1994) had students define the marking schedule for peer assessed experimental laboratory reports, and reported learning gains from the overall process. Catterall

(1995) had multiple choice and short essay tests peer marked by 120 marketing students. Learning gains from peer assessment were reported by 88% of participants, and impact on the ability to self-assess was reported by 76%. Hughes (1995) had first year pharmacology students use a detailed model-marking schedule. Their subsequent performance in practical work increased in comparison to previous years, whose ability on entry was identical. Segers and Dochy (2001) found no evidence of any effect of peer marking on learning outcomes.

3.5.2 Peer Assessment of Writing

In a business communication class, Roberts (1985) compared peer assessment in groups of five with staff assessment. Pre- and post-tests showed a statistically significant difference in favour of the peer condition. Falchikov (1986) involved 48 biological science students in discussion and development of essay assessment criteria. They felt the peer assessment process was difficult and challenging, but helped develop critical thinking. A majority reported increased learning and better self-organisation, while noting that it was time-consuming. The effects of teacher feedback, peer feedback and self-assessment were compared by Birkeland (1986) with 76 technicians, but no significant differences were found between conditions on test gains in paragraph writing ability. Richer (1992) compared the effects of peer group discussion of essays with teacher discussion and feedback. Grading of 174 pre- and post-test essays from 87 first year students indicated greater gains in writing proficiency in the peer feedback group. Hughes (1995) compared teacher, peer and self-assessment of written recording of pharmacology practical work, finding them equally effective.

Graner (1985) compared the effect of peer assessment and feedback in small groups to that of assessment of another's work alone using an editorial checklist. Both groups then rewrote their essays, and final grading was by staff. Both groups significantly improved from initial to final draft, and no significant difference was found between the groups. This suggests that practising critical evaluation can have generalised effects on the evaluator's own work, even in the absence of any external feedback about their own work. Chaudron (1983) compared the effectiveness of teacher feedback with feedback from peers with either English or another language as their first language. Students in all conditions showed a similar pattern of improvement. Working with 81 college students of ESL in Thailand and Hawaii, Jacobs and Zhang (1989) compared teacher, peer and self-assessment of essays. The type of assessment did not affect informational or rhetorical accuracy, but teacher and peer feedback was found to be more effective for grammatical accuracy.

3.5.3 Peer Assessment of Oral & Presentation Skills

Heun (1969) compared the effect on student self-concept of peer and staff assessment of four public speeches given by students. Compared to a control group, peer influence on the self-concept of students reached a significant level for the final speech, while instructor influence was non-significant across all four speeches. Mitchell and Bakewell (1995) found peer review of oral presentation skills led to significantly improved performance. Williams (1995) used peer assessment of oral presentations of critical incident analysis in undergraduate clinical practice nursing. Participants felt learning was enhanced, and the experience relevant to peer appraisal skills in the future working setting.

3.5.4 Peer Assessment of Group Work & Projects

Peer assessment has been used to help with the differentiation of individual contributions to small group projects (Conway, Kember, Sivan, & Wu, 1993; Falchikov, 1993; Goldfinch, 1994; Mathews, 1994), but empirical research on effects is hard to find. In a study of psychology students (Falchikov, 1993), group members and the lecturer negotiated self and peer assessment checklists of group process behaviours. Task-oriented behaviours proved easier to rate reliably than pro-social group maintenance behaviours such as facilitating the inclusion of quieter group members. Abson (1994) had marketing research students working in self-selected tutor-less groups use a simple five point rating scale on four criteria (co-operation, ideas, effort, reliability). A case study of one group suggested peer assessment might have made students work harder. Strachan and Wilcox (1996) used peer and self-assessment of group work to cope with increased enrolment in a third-year course in microclimatology. Students found this fair, valuable, enjoyable, and helpful in developing transferable skills in research, collaboration and communication.

3.5.5 Peer Assessment of Professional Skills

Peer assessment of professional skills can take place within the institution and/or out on practical placements or internships. In the latter case it is an interesting parallel to "peer appraisal" between staff in the workplace. It has been used by medical schools (e.g., Arnold, Willoughby, Calkins, Gammon, & Eberhart, 1981; Burnett & Cavaye, 1980; McAuley & Henderson, 1984), in pre-service teacher training (e.g., Litwack, 1974; Reich, 1975), and for other professions. It has also been used in short practical laboratory sessions (e.g., Stefani, 1992). Application is also reported in more exotic areas, such

as applied brass jury performances (Bergee, 1993), and a range of other musical performance arts (Hunter & Russ, 1995). Lennon (1995) considered tutor, peer and self-assessments of the performance of second year physiotherapy students in practical simulations. Students rated the learning experience highly overall. Also in physiotherapy, Orr (1995) used peer assessment in role-play simulation triads. Participants rated the exercise positively, but felt some anxiety about it. Ramsey and colleagues (1996) studied peer assessment of professional performance for 187 medical interns. The process was acceptable to the subjects, and reliability adequate despite the use of self-chosen raters.

Franklin (1981) compared self, peer and expert observational assessment of teaching sessions with pre-service secondary science teachers. There were no differences between the groups in skill acquisition. A similar study by Turner (1981) yielded similar results. Yates (1982) used reciprocal paired peer feedback with fourteen special education student teachers, followed by self-monitoring. The focus was the acquisition and maintenance of the skill of giving specific praise to learning-disabled pupils. Peer feedback was effective in increasing student teachers' use of motivational praise, but not content-based praise. With self-monitoring rates of both kinds of praise were maintained. Lasater (1994) paired twelve student teachers to give feedback to each other during twelve lessons in a 5-week practicum placement. The participants reported the personal benefits to be improved self-confidence and reduced stress. The benefits to their teaching included creative brainstorming and fine-tuning of lessons, resulting in improved organisation, preparation, and delivery of lessons.

3.5.6 Computer-Assisted Peer Assessment

Wider availability of word processing and electronic mail has created opportunities for formative peer assessment in electronic draft prior to final submission, as well as distributed collaborative writing. For example, Downing and Brown (1997) describe the collaborative creation of hypertexts by psychology students, which were published in draft on the World Wide Web and peer reviewed via email. Rushton, Ramsey and Rada (1993) and Rada, Acquah, Baker, & Ramsey (1993) reported peer assessment in a collaborative hypermedia environment. Good correspondence with staff assessment was evident, but the majority of computer science students were sceptical and preferred teacher-based assessment. Brock (1993) compared feedback from computerised text analysis programmes and from peer assessment and tutoring for 48 ESL student writers in Hong Kong. Both groups showed significant growth in writing performance. However, peer interaction was rated higher for helpfulness in improving content, and peer

supported students wrote significantly more words in post-intervention essays.

3.6 Summary and Conclusions on Peer Assessment

The reliability and validity of teacher assessment is not high. That of peer assessment tends to be at least as high, and often higher. Reliability tends to be higher in relation to: the degree of advancement in the course, the nature of the product or performance assessed, the extent to which criteria have been discussed and negotiated, the nature of assessment instrumentation, the extent to which an aggregate judgement rather than detailed components are compared, the amount of scaffolding, practice, feedback and monitoring, and the contingencies associated with the assessment outcome. Irrespective of relatively high reliability, student acceptance is variable.

In schools, research on peer assessment has focused less on reliability and more on effects. Students as young as grade 4 (9 years old) and those with special educational needs or learning disabilities have been successfully involved. The evidence on the effectiveness of peer assessment in writing is substantial, particularly in the context of peer editing. Here, peer assessment seems to be at least as effective in formative terms as teacher assessment, and sometimes more effective. The research on peer assessment of other learning outputs in school is as yet sparse, but merits exploration. In higher education, there is some evidence of impact of peer assessment on learning, especially in writing, sometimes greater than that of teacher assessment. In other areas, such as oral presentations, group skills, and professional skills, evidence for effects on learning are more dependent on softer data such as student subjective perceptions. Computer assisted peer assessment shows considerable promise.

In all sectors, much further development and evaluation is needed, with improved methodological quality and fuller and more detailed reporting of studies.

4. SELF VS. PEER ASSESSMENT

In Higher Education, Burke (1969) found self-assessments unreliable and peer assessments more reliable. By contrast, Falchikov (1986) found self-assessments were more reliable than peer assessments. However, Stefani (1994) found peer assessment more reliable. Saavedra and Kwun (1993) found outstanding students were the most discriminating peer assessors, but their self-assessments were not particularly reliable. Shore, Shore and Thornton (1992) found construct and predictive validity stronger for peer

than for self-evaluations, and stronger for more easily observable dimensions than for those requiring inferential judgement. Furnham and Stringfield (1994) reported greater reliability in peer assessments by subordinates and superiors than in self-assessments. Wright (1995) found self-assessment generally yielded lower marks than peer assessment, but less so in a structured module than in a more open ended one. Lennon (1995) found a high correlation between peer assessments of a piece of work (0.85), but lesser correlations between self and peer assessment (0.61 - 0.64). However, correlations between tutor and self-assessment were even lower (0.21), and those between tutor and peer assessment modest (0.34 - 0.55). Self-assessment was associated with under-marking and bunching at the median.

In general, peer assessment seems likely to correlate more highly with professional assessment than does self-assessment, and self and peer assessments do not always correlate well. Of course, triangulation between highly correlated measures is in any event redundant, and the processes here are at least as important as the actual judgement.

5. SUMMARY AND CONCLUSION RE SELF ASSESSMENT AND PEER ASSESSMENT

Both self and peer assessment have been successfully deployed in elementary and high schools and in higher education, including with very young students and those with special educational needs or learning disabilities. The reliability and validity of assessments by professional teachers is often low. The reliability and validity of self-assessment tends to be a little lower and more variable, while the reliability and validity of peer assessment tends to be as high or higher. Self-assessment is often assumed to have meta-cognitive benefits. There is some hard evidence that it can result in improvements in the effectiveness and quality of learning, which are at least as good as gains from teacher assessment, especially in relation to writing. However, this evidence is still quite limited. There is more substantial hard evidence that peer assessment can result in improvements in the effectiveness and quality of learning, which is at least as good as gains from teacher assessment, especially in relation to writing. In other areas the evidence is softer. Of course, self and peer assessment are not dichotomous alternatives - one can lead to and inform the other. Both can offer valuable triangulation in the assessment process and both can have measurable formative effects on learning, given good quality implementation. Both need training and practice, arguably on neutral products or performances, before full implementation, which should feature monitoring and moderation. Much further development is needed, with improved implementation and

evaluation quality. The following evidence-based guidelines for implementation are drawn from the research literature reviewed.

Good quality of organisation is important for implementation integrity and consistent and productive outcomes. Important planning issues evident in the literature are:

1. Clarify Purpose, Rationale, Expectations and Acceptability with all Stakeholders
2. Involve Participants in Developing and Clarifying Assessment Criteria
3. Match Participants & Arrange Contact (PA only)
4. Provide Quality Training, Examples and Practice
5. Provide Guidelines, Checklists or other tangible Scaffolding
6. Specify Activities and Timescale
7. Monitor the Process and Coach
8. Compare Aggregated Ratings, not multiple components (PA only)
9. Moderate Reliability and Validity of Judgements
10. Evaluate and Give Feedback

REFERENCES

Abson, D. (1994). The effects of peer evaluation on the behaviour of undergraduate students working in tutorless groups. In H. C. Foot, C. J. Howe, A. Anderson, A. K. Tolmie, & D. A. Warden (Eds.), *Group and Interactive Learning* (1st ed., Vol. 1, pp. 153-158). Southampton & Boston: Computational Mechanics.

Arnold, L., Willoughby, L., Calkins, V., Gammon, L., & Eberhart, G. (1981). Use of peer evaluation in the assessment of medical students. *Journal of Medical Education, 56,* 35-42.

Bangert-Drowns, R. L., Kulik, J. A., Kulik, C. C., & Morgan, M. (1991). The instructional effects of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Barnett, J. E., & Hixon, J. E. (1997). Effects of grade level and subject on student test score predictions. *Journal of Educational Research, 90* (3), 170-174.

Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education, 41,* 19-27.

Bernadin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work.* Belmont, CA: Wadsworth.

Birenbaum, M., & Dochy, F. (Eds.). (1996). *Alternatives in assessment of achievement, learning processes and prior knowledge.* Boston: Kluwer Academic.

Birkeland, T. S. (1986). Variations of feedback related to teaching paragraph structure to technicians. *Dissertation Abstracts International, 47,* 4362.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5* (1), 7-74.

Blatchford, P. (1997). Pupils' self assessments of academic attainment at 7, 11 and 16 years: Effects of sex and ethnic group. *British Journal of Educational Psychology, 67,* 169-184.

Boersma, G. (1995). *Improving student self-evaluation through authentic assessment.* ERIC Document Reproduction Service No. ED 393 885.

Boud, D. (1989). The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education, 14* (1), 20-30.

Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education, 15* (1), 101-111.

Boud, D. (Ed.). (1995). *Enhancing learning through self-assessment.* London & Philadelphia: Kogan Page.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society, *Studies in Continuing Education, 22* (2), 151-167.

Boud, D., Cohen, R., & Sampson, J. (Eds.). (2001). *Peer learning in Higher Education: Learning from and with each other.* London & Philadelphia: Kogan Page.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18* (5), 529-49.

Bouton, K., & Tutty, G. (1975). The effect of peer-evaluated student compositions on writing improvement. *The English Record, 3,* 64-69.

Brock, M. N. (1993). A comparative study of computerized text analysis and peer tutoring as revision aids for ESL writers. *Dissertation Abstracts International, 54,* 912.

Brown, S., & Dove, P. (1991). *Self and Peer Assessment.* Birmingham: Standing Conference on Educational Development (SCED).

Brown, S., & Knight, P. (1994). *Assessing Learners in Higher Education.* London: Kogan Page.

Burnett, W., & Cavaye, G. (1980). Peer assessment by fifth year students of surgery. *Assessment in Higher Education, 5* (3), 273-287.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65,* 245-281.

Burke, R. J. (1969). Some preliminary data on the use of self-evaluations and peer ratings in assigning university course grades. *Journal of Educational Research, 62* (10), 444-448.

Byard, V. (1989). *Power Play: The Use and Abuse of Power Relationships in Peer Critiquing.* Paper presented to Annual meeting of the conference on college composition and communication, Seattle WA, March 16-18 1989.

Byrd, D. R. (1994). Peer editing: Common concerns and applications in the foreign language classroom. *Unterrichtspraxis, 27* (1), 119.

Califano, L. Z. (1987). Teacher and peer editing: Their effects on students' writing as measured by t-unit length, holistic scoring, and the attitudes of fifth and sixth grade students. *Dissertation Abstracts International. 49* (10), 2924.

Catterall, M. (1995). Peer learning research in marketing. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing student learning through peer tutoring in higher education: Section 3 - Implementing.* (1st ed., Vol. 1, pp. 54-62). Coleraine, NI: University of Ulster.

Chaudron, C. (1983). *Evaluating Writing: Effects of Feedback on Revision.* Paper presented at the Annual TESOL Convention (17th, Toronto, Ontario, March 16-19, 1983).

Cicchetti, D. V. (1982). On peer review - We have met the enemy and he is us. *Behavioral and Brain Sciences, 5* (2), 205-206.

Cicchetti, D. V. (1991). The reliability of peer-review for manuscript and grant submissions - A cross-disciplinary investigation. *Behavioral and Brain Sciences, 14* (1), 119-134.

Cohen, E. G. (1982). Expectation states and interracial interaction in school settings. *Annual Review of Sociology, 8,* 209-235.

Cole, D. A. (1991). Change in self-perceived competence as a function of peer and teacher evaluation. *Developmental Psychology, 27* (4), 682-688.

Conway, R., Kember, D., Sivan, A., & Wu, M. (1993). Peer assessment of an individual's contribution to a group project. *Assessment and Evaluation in Higher Education, 18*(1), 45-56.

Cover, B. T. L. (1987). Blue-pencil workers: the effects of a peer editing technique on students' editing skills and attitudes toward writing at the seventh grade level. *Dissertation Abstracts International, 48* (08), 1968.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58* (4), 438-481.

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Teaching International, 37* (4), 346-355.

Davis, J. K., & Rand, D. C. (1980). Self-grading versus instructor grading. *Journal of Educational Research, 73* (4), 207-11.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education: A review. *Studies in Higher Education, 24 (3),* 31 -350.

Downing, T., & Brown, I. (1997). Learning by cooperative publishing on the World Wide Web. *Active Learning, 7,* 14-16.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040-1047.

Dweck, C. S., & Bush, E. S. (1976). Sex differences in learned helplessness: Differential debilitation with peer and adult evaluators. *Developmental Psychology, 12* (2), 1.

El-Koumy, A. S. A. (2001). *Effects of Student Self-Assessment on Knowledge Achievement and Academic Thinking.* Paper presented at the Annual Meeting of the Integrated English Language Program-II (3rd, Cairo, Egypt, April 18-19, 2001). ERIC Document Reproduction Service No. ED 452 731.

Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education, 11* (2), 146-166.

Falchikov, N. (1993). Group process analysis - Self and peer assessment of working together in a group. *Educational & Training Technology International, 30 (3)*, 275-284.

Falchikov, N. (1995). Peer feedback marking - Developing peer assessment. *Innovations in Education and Training International, 32,* 175-187.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education.* London & New York: RoutledgeFalmer.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59* (4), 395-430.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in Higher Education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70* (3), 287-322.

Farh, J., Cannella, A. A., & Bedeian, A. G. (1991). Peer ratings: The impact of purpose on rating quality and user acceptance. *Group and Organization Studies, 16,* 367-386.

Fedor, D. B., & Bettenhausen, K. L. (1989). The impact of purpose, participant preconceptions, and rating level on the acceptance of peer evaluations. *Group and Organization Studies, 14,* 182-197.

Fernandes, M., & Fontana, D. (1996). Changes in control beliefs in Portuguese primary school pupils as a consequence of the employment of self-assessment strategies. *British Journal of Educational Psychology, 66,* 301-313.

Fontana, D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary-school pupils. *British Journal of Educational Psychology, 64,* 407-417.

Franklin, C. A. (1981). Instructor versus peer feedback in microteaching on the acquisition of confrontation; illustrating, analogies, and use of examples; and question-asking teaching skills for pre-service science teachers. *Dissertation Abstracts International, 42,* 3565.

Freedman, S. W. (1992). Outside-in and inside-out: Peer response groups in two ninth grade classes. *Research in the Teaching of English, 26,* 71-107.

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education, 20,* 289-300.

Fry, S. A. (1990). Implementation and evaluation of peer marking in higher education. *Assessment and Evaluation in Higher Education, 15,* 177-189.

Frye, A. W., Richards, B. F., Bradley, E. W., & Philp, J. R. (1992). The consistency of students self-assessments in short-essay subject-matter examinations. *Medical Education, 26* (4), 310-316.

Furnham, A., & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology, 67,* 57-67.

Gaier, E. L. (1961). Student self assessment of final course grades. *Journal of Genetic Psychology, 98* (1), 63-67.

Gaillet, L. I. (1992). *A foreshadowing of modern theories and practices of collaborative learning: The work of the Scottish rhetorician George Jardine.* Paper presented at the 43rd Annual Meeting of the Conference on College Composition and Communication, Cincinnati OH, March 19-21 1992.

Gere, A. R., & Abbot, R. D. (1985). Talking about writing: The language of writing groups. *Research in the Teaching of English, 19,* 362-379.

Goldfinch, J. (1994). Further developments in peer assessment of group projects. *Assessment and Evaluation in Higher Education, 19* (1), 29-35.

Graner, M. H. (1985). Revision techniques: Peer editing and the revision workshop. *Dissertation Abstracts International, 47,* 109.

Griffee, D. T. (1995). *A longitudinal study of student feedback: Self-assessment, Course Evaluation and Teacher Evaluation.* Longman: Birmingham, Alabama.

Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology, 20* (1), 28- 32.

Hendrickson, J. M., Brady, M. P., & Algozzine, B. (1987). Peer-mediated testing: The effects of an alternative testing procedure in higher education. *Educational and Psychological Research, 7* (2), 91-102.

Henry, S. E. (1979). Sex and locus of control as determinants of children's responses to peer versus adult praise. *Journal of Educational Psychology, 71* (5), 605.

Herman, J., Gearhart, M., & Baker, E. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment, 1* (3), 201-224.

Heun, L. R. (1969). *Speech Rating as Self-Evaluative Behavior: Insight and the Influence of Others.* PhD Dissertation, Southern Illinois University.

Heywood, J. (1988). *Assessment in higher education.* Chichester: John Wiley.

Holley, C. A. B. (1990). The effects of peer editing as an instructional method on the writing proficiency of selected high school students in Alabama. *Dissertation Abstracts International, 51* (09), 2970.

Hounsell, D., & McCulloch, M. (1999). Assessing skills in Scottish higher education. In E. Dunne (Ed.) *The learning society: International perspectives on core skills in higher education* (pp.149-158). London: Kogan Page.

Hughes, I. E. (1995). Peer assessment. *Capability, 1* (3), 39-43.

Hughes, I. E., & Large, B. J. (1993a). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education, 18,* 379-385.

Hughes, I. E., & Large, B. J. (1993b). Assessment of students' oral communication skills by staff and peer groups. *New Academic, 2* (3), 10-12.

Hunter, D., & Russ, M. (1995). Peer assessment in performance studies. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing Student Learning Through Peer Tutoring in Higher Education: Section 3 - Implementing.* (1st ed., Vol. 1, pp. 63-65). Coleraine, NI: University of Ulster.

Jacobs, G., & Zhang, S. (1989). *Peer Feedback in Second Language Writing Instruction: Boon or Bane?* Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).

Joines, S. M. B., & Sommerich, C. M. (2001). Comparison of self-assessment and partnered-assessment as cost-effective alternative methods for office workstation evaluation. *International Journal of Industrial Ergonomics, 28* (6), 327-340.

Karegianes, M. L., Pascarella, E. T., & Pflaum, S. W. (1980). The effects of peer editing on the writing proficiency of low-achieving tenth grade students. *Journal of Educational Research, 73* (4), 203-207.

Kaye, M. M., & Dyason, M. D. (1999). Achieving a competitive focus through self-assessment. *Total Quality Management, 10* (3), 373-390.

Koretz, D., Stecher, B., Klein, S., & Mc Caffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement, 13* (3), 5-16.

Korman, M., & Stubblefield, R. L. (1971). Medical school evaluation and internship performance. *Journal of Medical Education, 46,* 670-673.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1,* 279-308.

Lagana, J. R. (1972). *The development, implementation and evaluation of a model for teaching composition which utilizes individualized learning and peer grouping.* Unpublished doctoral thesis, University of Pittsburgh, Pittsburgh, PA.

Lasater, C. A. (1994). Observation feedback and analysis of teaching practice: Case studies of early childhood student teachers as peer tutors during a preservice teaching practicum. *Dissertation Abstracts International, 55,* 1916.

Lawrence, M. J. (1996). *The effects of providing feedback on the characteristics of student responses to a videotaped high school physics assessment.* Unpublished doctoral thesis, Rutgers University, New Brunswick, NJ.

Lee, B. (1999). *Self-assessment for pupils with learning difficulties.* Slough UK: National Foundation for Educational Research.

Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education, 26* (1), 61-72.

LeMahieu, P., Gitomer, D. H., & Eresh, J. T. (1995). Portfolios in large-scale assessment: Difficult but not impossible. *Educational Measurement, 14* (3), 11-16, 25-28.

Lennon, S. (1995). Correlations between tutor, peer and self assessments of second year physiotherapy students in movement studies. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing Student Learning Through Peer Tutoring in Higher Education: Section 3 - Implementing.* (1st ed., Vol. 1, pp. 66-71). Coleraine, NI: University of Ulster.

Litwack, M. (1974). A study of the effects of authority feedback, peer feedback, and self feedback on learning selected indirect-influence teaching skills. *Dissertation Abstracts International, 35,* 5762.

Lloyd, J. (1982). Reactive effects of self-assessment and self-recording on attention to task and academic productivity. *Learning Disability Quarterly, 5* (3), 216-27.

Longhurst, N., & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation, 23* (4), 319-330.

MacArthur, C. A., Schwartz, S. S., & Graham, S. (1991). Effects of a reciprocal peer revision strategy in special education classrooms. *Learning Disabilities Research and Practice, 6* (4), 201-210.

MacKenzie, L. (2000). Occupational therapy students as peer assessors in viva examinations. *Assessment & Evaluation in Higher Education, 25* (2), 135- 147.

MacLellan, E. (2001). Assessment for learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education, 26* (4), 307-318.

Magin, D. J. (200 la). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education, 26* (1), 53-63.

Magin, D. J. (2001b). A novel technique for comparing the reliability of multiple peer assessments with that of single teacher assessments of group process work. *Assessment and Evaluation in Higher Education, 26* (2), 139-152.

Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies In Higher Education, 26* (3), 287-298.

Marcoulides, G. A., & Simkin, M. G. (1991). Evaluating student papers: The case for peer review. *Journal of Education for Business, 67,* 80-83.

Marcoulides, G. A., & Simkin, M. G. (1995). The consistency of peer review in student writing projects. *Journal of Education for Business, 70* (4), 220-223.

Marienau, C. (1999). Selfassessment at work: Outcomes of adult learners' reflections on practice. *Adult Education Quarterly, 49* (3), 135-146.

Mathews, B. P. (1994). Assessing individual contributions - Experience of peer evaluation in major group projects. *British Journal of Educational Technology, 25,* 19-28.

McAuley, R. G., & Henderson, H. W. (1984). Results of the peer assessment program of the college of physicians and surgeons of Ontario. *Canadian Medical Association Journal, 131*, 557-561.

McCurdy, B. L., & Shapiro, E. S. (1992). A comparison of teacher-monitoring, peer-monitoring, and self-monitoring with curriculum-based measurement in reading among students with learning-disabilities. *Journal of Special Education, 26* (2), 162-180.

McDonald, B. (2002). *Self assessment and academic achievement.* Unpublished Ph.D. thesis, University of the West Indies, Cave Hill, Barbados, West Indies.

Midgley, D. F., & Petty, M. (1983). *Final report on the survey of graduate opinions on general education.* Kensington: University of New South Wales.

Miller, M. (1988). *Self-Assessment in students with learning handicaps.* Paper presented at the Annual Convention of the Council for Exceptional Children (66th, Washington, DC, March 28-April 1, 1988). ERIC Document Reproduction Service No. EC 210 383.

Mills, L. (1994). *Yes, It Can Work!: Portfolio Assessment with Preschoolers.* Paper presented at the Association for Childhood Education International Study Conference (New Orleans, LA, March 30-April 2, 1994). ERIC Document Reproduction Service No. ED 372 857.

Mitchell, V. W., & Bakewell, C. (1995). Learning without doing - Enhancing oral presentation skills through peer-review. *Management Learning, 26,* 353-366.

Mockford, C. D. (1994). The use of peer group review in the assessment of project work in higher education. *Mentoring and Tutoring, 2* (2), 45-52.

Newstead, S. E. (1996). The psychology of student assessment. *The Psychologist, 9,* 543-7.

Newstead, S., & Dennis, I. (1994). Examiners examined. *The Psychologist, 7,* 216-9.

Ney, J. W. (1989). *Teacher-Student Cooperative Learning in the Freshman Writing Course.* ERIC Document Reproduction Service.

Ninness, H. A. C., Ellis, J., & Ninness, S. K. (1999). Self-Assessment as a learned reinforcer during computer interactive math performance - An Experimental analysis. *Behavior Modification, 23* (3), 403-418.

Ninness, H. A. C., Ninness, S. K., Sherman, S., & Schotta, C. (1998). Argumenting computer-interactive self-assessment with and without feedback. *Psychological Record, 48* (4), 601-616.

O'Donnell, A. M., & Topping, K. J. (1998). Peers assessing peers: Possibilities and problems. In K. J. Topping, & S. W. Ehly (Eds.), *Peer assisted learning* (Chapter 14, pp. 255-278). Mahwah NJ: Lawrence Erlbaum.

Olson, V. L. B. (1986). The effects of revision instruction and peer response groups on the revision behaviors, quality of writing and attitude toward writing of sixth grade students. *Dissertation Abstracts International, 47* (12), 4310.

Olson, V. L. B. (1990). The revising processes of sixth-grade writers with and without peer feedback. *Journal of Educational Research, 84* (1), 1.

Orpen, C. (1982). Student vs lecturer assessment of learning: A research note. *Higher Education, 11,* 567-572.

Orr, A. (1995). Peer assessment in a practical component of physiotherapy education. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing Student Learning Through Peer Tutoring in Higher Education: Section 3 - Implementing. (*1st ed., Vol. 1, pp. 72-78). Coleraine, NI: University of Ulster.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 25* (1), 23-38.

Paris, S. G., & Newman, R. S. (1990). Developmental aspects of self-regulated learning. *Educational Psychologist, 25,* 87-102.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36* (2), 89-101.

Paulson, E. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48* (5), 60-63.

Pierson, H. (1967). *Peer and teacher correction: A comparison of the effects of two methods of teaching composition in grade 9 English classes.* Unpublished doctoral thesis, New York University, New York.

Pond, K., Ul-Haq, R., & Wade, W. (1995). Peer review: A precursor to peer assessment. *Innovations in Education and Training International, 32,* 314-323.

Rada, R., Acquah, S., Baker, B., & Ramsey P. (1993). Collaborative learning and the MUCH System. *Computers and Education, 20,* 225-233.

Ramsey, P. G., Carline, J. D., Blank, L. L., & Wenrich M. D. (1996). Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Academic Medicine, 71,* 364-370.

Raphael, T. E. (1986). *The impact of text structure instruction and social context on students' comprehension and production of expository text.* East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Reich, R. (1975). The effect of peer feedback on the use of specific praise in student-teaching. *Dissertation Abstracts International, 37,* 925.

Richer, D. L. (1992). The effects of two feedback systems on first year college students' writing proficiency. *Dissertation Abstracts International, 53,* 2722.

Rijlaarsdam, G. (1987). *Effects of peer evaluation on writing performance, writing processes, and psychological variables.* Paper presented at the 38th Annual Meeting of the

Conference on College Composition and Communication, Atlanta GA, March 19-21, 1987.

Rijlaarsdam, G., & Schoonen, R. (1988). *Effects of a teaching program based on peer evaluation on written composition and some variables related to writing apprehension.* Amsterdam: Stichting Centrum voor Onderwijsonderzoek, Amsterdam University.

Riley, S. M. (1995). Peer responses in an ESL writing class: Student interaction and subsequent draft revision. *Dissertation Abstracts International, 56,* 3031.

Roberts, W. H. (1985). The effects of grammar reviews and peer-editing on selected collegiate students' ability to write business letters effectively. *Dissertation Abstracts International, 47,* 1994.

Rocklin, T. R., O'Donnell, A. M., & Hoist, P. M (1995). Effects and underlying mechanisms of self-adapted testing. *Journal of Educational Psychology, 87,* 103-116.

Rogers, C. R. (1983). *Freedom to learn for the 80s.* Columbus OH: Charles E. Merrill.

Ross, J. A. (1995). Effects of feedback on student behavior in cooperative learning groups in a grade-7 math class. *Elementary School Journal, 96* (2), 125-143.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15* (1), 1-20.

Rowntree, D. (1977). *Assessing students: How shall we know them?* London: Harper & Row.

Rudd, T. J., & Gunstone, R. F. (1993). *Developing self-assessment skills in grade 3 science and technology: The importance of longitudinal studies of learning.* Paper presented at the Annual Meetings of the National Association for Research in Science Teaching (Atlanta, GA, April 15-18, 1993) and the American Educational Research Association (Atlanta, GA, April 12-16, 1993). ERIC Document Reproduction No. ED 358 103.

Rushton, C., Ramsey, P., & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment - A case-study. *Journal of Computer-Based Instruction, 20,* 75-80.

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78,* 450-462.

Salend, S. J., Whittaker, C. R., & Reeder, E. (1993). Group evaluation - a collaborative, peer-mediated behavior management system. *Exceptional Children, 59* (3), 203-209.

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13,* 89-99.

Samway, K. D. (1993). This is hard, isn't it - children evaluating writing. *Tesol Quarterly, 27* (2), 233-257.

Schunk, D. H. (1996). L*earning theories: An educational perspective* (2nd ed.). Englewood Cliffs NJ: Prentice-Hall.

Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: The value-added of the students' perspective. *Studies in Higher Education, 26* (3), 327-343.

Shore, T. H., Shore, L. M., & Thornton, G. C. (1992). Construct validity of self evaluations and peer evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology, 77,* 42-54.

Sink, C. A., Barnett, J. E., & Hixon, J. E. (1991). Self-regulated learning and achievement by middle-school children. *Psychological Reports, 69* (3), 979-989.

Sobral, D. T. (1997). Improving learning skills: A self-help group approach. *Higher Education, 33* (1), 39-50.

Stefani, L. A. J. (1992). Comparison of collaborative self, peer and tutor assessment in a biochemistry practical. *Biochemical Education, 20* (3), 148-151.

Stefani, L. A. J. (1994). Peer, self and tutor assessment - Relative reliabilities. *Studies in Higher Education, 19* (1), 69-75.

Stoddard, B., & MacArthur, C. A. (1993). A peer editor strategy - guiding learning-disabled students in response and revision. *Research in the Teaching of English, 27*(1), 76-103.

Strachan, I. B., & Wilcox, S. (1996). Peer and self assessment of group work: Developing an effective response to increased enrolment in a third-year course in microclimatology. *Journal of Geography in Higher Education, 20* (3), 343- 353.

Supovitz, J. A., MacGowan, A., & Slattery J. (1997). Assessing agreement: An examination of the interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment, 4* (3), 237-259.

Taylor, I. (1995). Understanding computer software: Using peer tutoring in the development of understanding of some aspects of computer software. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing Student Learning Through Peer Tutoring in Higher Education: Section 3 - Implementing.* (1st ed., Vol. 1, pp. 87-89). Coleraine, NI: University of Ulster.

Topping, K. J. (1996a). The effectiveness of peer tutoring infurther and higher education: A typology and review of the literature. *Higher Education, 32* (3), 321 -345. (Also in S. Goodlad. (Ed.). (1998). *Mentoring and tutoring by students.* London & Stirling VA: Kogan Page.)

Topping, K. J. (1996b). *Effective peer tutoring in further and higher education* (SEDA Paper 95). Birmingham: Staff and Educational Development Association.

Topping, K. J. (1998). Peer assessment between students in college and university. *Review of Educational Research, 68* (3), 249-276.

Topping, K. J. (1999). Formative assessment of reading comprehension by computer: Advantages and disadvantages of the Accelerated Reader software. *Reading OnLine* (I.R.A.) [Online]. Available www.readingonline.org/critical/topping/ [November 4]. (hypermedia).

Topping, K. J. (2001a). *Peer assisted learning: A practical guidefor teachers.* Cambridge MA: Brookline Books.

Topping, K. J. (2001b). *Tutoring by peers, family and volunteers.* Geneva: International Bureau of Education, United Nations Educational, Scientific and Cultural Organisation (UNESCO). [Online] Available: www.ibe.unesco.org/lnternational/Publicalions/EducationalPractices/prachomc.htm [January 1] (Also in translation in Chinese and Spanish).

Topping, K. J., & Ehly, S. W. (Eds.). (1998). *Peer-assisted learning.* Mahwah NJ & London UK: Lawrence Erlbaum.

Topping, K. J., & Ehly, S. W. (2001). Peer assisted learning: A framework for consultation. *Journal of Educational and Psychological Consultation, 12* (2), 113-132.

Topping, K. J., & Sanders, W. L. (2000). Teacher effectiveness and computer assessment of reading: Relating value added and learning information system data. *School Effectiveness and School Improvement, 11* (3), 305-337.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education, 25* (2), 149-169.

Towler, L, & Broadfoot, P. (1992). Self-assessment in the primary school. *Educational Review, 44* (2), 137-151.

Turner, R. F. (1981). The effects of feedback on the teaching performance of preservice teachers involved in a microteaching experience. *Dissertation Abstracts International, 42,* 3116.

Wade, L. K. (1988). An analysis of the effects of a peer feedback procedure on the writing behavior of sixth-grade students. *Dissertation Abstracts International, 50* (05), 2181.

Ward, M, Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education, 7* (1), 63-80.

Wassef, A., Mason, G., Collins, M. L., O'Boyle, M., & Ingham, D. (1996). In search of effective programs to address students' emotional distress and behavioral problems: Student assessment of school-based support groups. *Adolescence, 31* (12), 1-16.

Weaver, M. E. (1995). Using peer response in the classroom: Students' perspectives. *Research and Teaching in Developmental Education, 12,* 31-37.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 13-40.

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31,* 369-395.

Weeks, J. O., & White, M. B. (1982). *Peer editing versus teacher editing: Does it make a difference?.* Paper presented at Meeting of the North Carolina Council of the International Reading Association, Charlotte NC, March 7-9, 1982.

Williams, J. (1995). Using peer assessment to enhance professional capability. In M. Yorke (Ed.), *Assessing Capability in Degree and Diploma Programmes.* (1st ed., Vol. 1, pp. 59-67). Liverpool: Centre for Higher Education Development, Liverpool John Moores University.

Wolfe, L., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8,* 227-242.

Wright, L. (1995). All students will take more responsibility for their own learning. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing Student Learning Through Peer Tutoring in Higher Education: Section 3 - Implementing.* (1st ed., Vol. 1, pp. 90-92). Coleraine, NI: University of Ulster.

Yates, J. A. (1982). The effects of peer feedback and self-monitoring on student teachers' use of specific praise. *Dissertation Abstracts International, 43,* 2321.

Zoller, Z., & Ben-Chaim, D. (1997). *Student self-assessment in Science Examinations: Is it compatible with that of teachers?* Paper presented at the meeting of the European Association for Research on Learning and Instruction, Greece, Athens, August 26-30.