

Peer Assessment between Students in Colleges and Universities

Author(s): Keith Topping

Source: *Review of Educational Research*, Vol. 68, No. 3 (Autumn, 1998), pp. 249-276

Published by: American Educational Research Association

Stable URL: <https://www.jstor.org/stable/1170598>

Accessed: 06-05-2019 22:36 UTC

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/1170598?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

American Educational Research Association is collaborating with JSTOR to digitize, preserve and extend access to *Review of Educational Research*

Peer Assessment Between Students in Colleges and Universities

Keith Topping
University of Dundee

A definition and typology of peer assessment between students in higher education is proposed, and the theoretical underpinnings of the method are discussed. A review of the developing literature follows, including both process and outcome studies. This indicates that peer assessment is of adequate reliability and validity in a wide variety of applications. Peer assessment of writing and peer assessment using marks, grades, and tests have shown positive formative effects on student achievement and attitudes. These effects are as good as or better than the effects of teacher assessment. Evidence for such effects from other types of peer assessment (of presentation skills, group work or projects, and professional skills) is, as yet, more limited. Computer-assisted peer assessment is an emerging growth area. Important factors in successful implementation are summarized, and recommendations for future research and practice are made.

In the increasingly diverse context of higher education, the formative, heuristic purposes of assessment have become more prominent. Formative assessment aims to improve learning while it is happening in order to maximize success rather than merely determine success or failure only after the event. Such 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; Brown & Knight, 1994). Given this emphasis, interest has grown in self-assessment by students (Boud & Holmes, 1981; Boud, Churches, & Smith, 1986; Gale, 1984) and in peer assessment, which share common features.

Purpose and Structure of This Review

It is surprising that while the literature on self-assessment has been reviewed (Boud & Falchikov, 1989; Falchikov & Boud, 1989), that on peer assessment has not. The present review is a first attempt at filling this gap. Its objectives are several: to determine the extent, nature, and quality of the literature to date; to develop a typology of peer assessment; to explore the theoretical underpinnings of peer assessment and elucidate the mechanisms through which it might have its effects; and to outline directions for future research and practice. Evidence in the literature about peer assessment through marks, grades, and tests is considered,

This project was carried out with support from the Scottish Higher Education Funding Council.

and reliability and validity are discussed. More detailed formative peer assessments of oral presentations, written outputs, group work and projects, and professional skills are then reviewed. Developments in computer-assisted peer assessment are noted. Important factors in successful implementation emerging from the literature are summarized with a view to future replications. The essential question is, What types of peer assessment are in use in higher education, and to what effect?

Methodology for the Review

The *Social Science Citation Index*, the *Educational Resources Information Center (ERIC)*, and *Dissertation Abstracts International* were searched on-line for the years 1980 to 1996. Search keywords included the following: *peer assessment*, *peer marking*, *peer correction*, *peer rating*, *peer feedback*, *peer review*, and *peer appraisal* (together with *university*, *college*, and *higher education*). *Peer review* yielded many items concerned with peer review of academic writing and of professional (usually medical) facilities. *Peer appraisal* yielded many studies of hierarchical or lateral peer appraisal of work skills in professional employment. A manual search of references in the full text of retrieved items yielded further relevant studies published before 1980.

All items unequivocally focusing on peer assessment between students in higher education were selected for inclusion, 109 in all. Forty-two articles were considered purely descriptive and anecdotal, while 67 (62%) included outcome data gathered in an orderly research process. Of the latter, studies of higher methodological quality are highlighted and discussed at greater length later. These include those featuring more detailed and rigorous analysis of process (e.g., Falchikov, 1995a), measures of known reliability and validity (e.g., Haaga, 1993), and/or quasi-experimental investigation of outcomes (e.g., Heun, 1969). However, at this early stage of development of the field, such studies are not numerous. A definition and typology of peer assessment were developed from immersion in this literature and with reference to similar typologies of other forms of peer-assisted learning (Topping, 1996; Topping & Ehly, 1998).

Definition and Typology of Peer Assessment

Definition of Peer Assessment

This review is concerned only with peer assessment between students in higher education of similar degree status, usually in the same course of study and often in the same year. It excludes the practice of paying postgraduates to grade the work of undergraduates, thereby acting as surrogate staff members. In this article, *peer assessment* is defined as an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status. The varying nomenclature adopted by different authors in the literature can prove confusing and needs careful scrutiny.

Elements of a Typology

It is evident from the literature that peer assessment activities in higher education vary widely. Thus, sweeping conclusions about peer assessment in general

are unlikely to be meaningful, irrespective of issues of implementation quality. Some of the main parameters of variation between projects reported in the literature are described subsequently and summarized in Table 1.

- (1) Studies were located in many different curriculum areas or subjects, suggesting that peer assessment is potentially applicable to virtually all areas.
- (2) The objectives specified for or implicit in projects varied in number and type. For example, some projects aimed to save staff assessment time or other costs (often when confronted with greatly enlarged classes), while other projects aimed to add value in terms of cognitive, metacognitive, or other gains for participants.
- (3) Allied to the preceding, a general purpose or focus could often be identified: a summative orientation, a formative orientation, or both.
- (4) A wide range of products or outputs were subjected to peer assessment, including test performance involving scoring and grades. Marks or grades were also applied by peers to products such as writing or presentations. Detailed open-ended assessment and feedback were more frequently applied to continuous writing, oral/audiovisual presentations, group work projects, and other skilled professional behaviors.
- (5) The relationship of the peer assessment to "official" staff assessment varied. In some projects, the previous staff assessment continued unchanged, and the peer assessment was clearly supplementary (often intended to formatively add value). In other projects, the peer assessment functioned as a substitute for part or all of the previous staff assessment (although in the latter case, quality assurance checks were usually still made by staff on a sampling basis).
- (6) Associated with the preceding was whether the peer assessment contributed to the assessee's yearly or overall official degree grade or grade point average.
- (7) The directionality of peer assessment also varied. It could be unidirectional (assessor to assessee), reciprocal, or mutual.
- (8) Even in cases in which mutual peer assessment operated, anonymous assessment was still possible, the assessee remaining unaware of the source of any particular peer assessment.
- (9) Associated with the preceding was variation in degree of personal contact. For example, when assessing written products, personal contact between assessor and assessee is not necessary, and feedback can be written (including by e-mail).
- (10) Year of study between assessor and assessee also varied. Although most peer assessment occurred between students in the same year of study, some studies of peer assessment of professional skills and behaviors cut across years. In courses with many mature students, the ages and life experiences of the participants could prove very different, even in a "same year" project. The few cross-year studies were likely to place the more advanced students in the role of assessor and to involve unidirectional assessment.
- (11) Even within the same year of study, students can be matched to ensure an ability differential so that the more expert students assess those less expert. Such an arrangement was little reported, however. Indeed, little attention was paid in the literature to the relative ability of the assessor and the person assessed, although presumably this might need to be controlled for

Topping

TABLE 1

A Typology of Peer Assessment in Higher Education

Variable	Range of Variation
1 Curriculum area/subject	All
2 Objectives	Of staff and/or students? Time saving or cognitive/affective gains?
3 Focus	Quantitative/summative or qualitative/formative or both?
4 Product/output	Tests/marks/grades or writing or oral presentations or other skilled behaviours?
5 Relation to staff assessment	Substitutional or supplementary?
6 Official weight	Contributing to assessee final official grade or not?
7 Directionality	One-way, reciprocal, mutual?
8 Privacy	Anonymous/confidential/public?
9 Contact	Distance or face to face?
10 Year	Same or cross year of study?
11 Ability	Same or cross ability?
12 Constellation Assessors	Individuals or pairs or groups?
13 Constellation Assessed	Individuals or pairs or groups?
14 Place	In/out of class?
15 Time	Class time/free time/informally?
16 Requirement	Compulsory or voluntary for assessors/ees?
17 Reward	Course credit or other incentives or reinforcement for participation?

maximum benefit.

- (12) Constellations of assignment of assessors to assessees varied. Although one assessor to one assessee was the modal constellation, both assessors and assessees could be matched to individuals, pairs, or groups.
- (13) Places and times for peer assessment activities varied. Most took place during formal class time, but some occurred out of class time and were coordinated and accounted for only when the class met.
- (14) Most peer assessment activities appeared to be required by staff rather than being voluntary.
- (15) A few projects awarded a modest amount of course credit for participation as an assessor. Other extrinsic reward or reinforcement was unusual.

Illustrative Case Study

The study of Falchikov (1995a) is summarized here. The objectives of the project were to improve the quality of the learning process, sharpen critical abilities in students, and increase student autonomy. The subjects were 13 human developmental psychology students (12 women and 1 man; mean age approximately 21 years) in the third year of a 4-year undergraduate course in biological sciences.

First, students carried out an individual exercise. They visited the library, selected from any relevant journal an experimental study on any topic in preadolescent development of interest to them, summarized it, and suggested what experiment might be fruitfully carried out next. They were aware they would be required to make a 10-minute oral presentation to the group on the basis of this, using visual aids and handouts as appropriate. They were then told they would be required to conduct a peer assessment of the presentations and asked to consider the qualities of good and bad presentations. From this, a composite checklist of qualitative assessment criteria was developed. Each student was also asked to identify the best and weakest feature of each presentation. Presentations were also to be awarded a mark out of 20, and the aggregate of these marks was to carry equal weight with the staff assessment of each presentation toward the coursework grade for the presenter.

Peer marking was conducted anonymously, but detailed peer feedback was given orally to the presenters by both students and staff. The students also completed an evaluative questionnaire about the peer assessment exercise. Aggregate peer marks were very similar to staff marks, "overmarking" being slightly more frequent than "undermarking." Detailed feedback on the agreed criteria was both global and specific (relating to the whole presentation or part of it). Much less agreement about the strongest and weakest features of presentations was evident, but there was less variety and thus more agreement about weaknesses.

The students' subjective responses to the procedure emphasized its fairness (assessment by a greater number of people) and the formative utility of detailed feedback. Least-liked features included social embarrassment (particularly with respect to identifying weaknesses) and the cognitive challenge (and strain) of the exercise. Traditional staff assessment was characterized as less informative, less effortful for students, and more "accurate." It was concluded that the objectives of the project were largely met. The author noted that social embarrassment might prove to be a greater problem in small groups of long standing than in large, newly constituted groups. This was the first experience of peer assessment for these students, and further experience might improve acceptance and stabilize reliability of marking even further. This study is of interest for its combination of "quantitative" and "qualitative" peer assessment. The use of aggregate peer marks obviously defends against the impact of singular rogue scorers.

In terms of the typology of peer assessment, this project was an example of a same-year, bilaterally quantitative and qualitative, in-class, compulsory, supplementary, mutual group peer assessment system in oral presentation skills in psychology. It was targeted on cognitive and affective gains, contributing to official grade, with quantitative aspects anonymous and qualitative aspects public and face to face without extrinsic reinforcement. Since all participants assessed all other participants, matching by ability did not occur. The project illuminated some of the psychological processes and mechanisms operating.

Theorizing About Peer Assessment

Problems

Given the many different types of peer assessment, establishing a single overarching theory or model of the process seems likely to be difficult. It is rendered even more difficult by the origins of the literature in a multiplicity of subject specialties with very different theoretical perspectives or a reliance on “commonsense” interpretations. As Patterson (1996) has pointed out, the curriculum paradigm, objectives, and ethos are likely to inform perceptions of the purpose of peer assessment. Beyond this, Fry (1990) noted that Aptitude Treatment interactions can be expected and that what are construed as advantages and disadvantages will vary according to the values, objectives, and capabilities of each participant.

One might expect a theory of peer assessment to draw on social constructivism—the joint construction of knowledge through discourse and other interactivity—even when assessor and assessee have no face-to-face contact. Communication and social skills seem to be implicit. The need to communicate the assessment to another should create purpose and accountability, the language for this purpose both leading and following the assessor’s internal thought processes, as proposed by Vygotsky (1978). The Vygotskian concept of scaffolded learning (partially supported by a more competent other) might also be involved. This would presumably depend on whether the peer assessor merely identified weaknesses in the assessed work or also identified strengths or suggested how the work could be improved.

Reciprocal same-ability peer assessment, between partners who are equally but differently competent, seems to fit better into the Piagetian model of cognitive conflict. This has relevance to cooperative work groups of similar ability but heterogeneous opinions. The Vygotskian, Piagetian, and other theoretical paradigms relevant to various forms of peer-assisted learning are discussed further in Topping and Ehly (1998). In fact, the literature features many hypotheses about the mechanisms through which peer assessment might create its effects, although direct tests of their validity are still scarce. They are discussed here by domain: cognition and metacognition, affect, social and transferable skills, and systemic benefits.

Cognition and Meta-Cognition

For the assessor. Peer assessment is reflexive. The expression *learning by teaching*, frequently applied to peer tutoring, might become *learning by assessing* in the current context. Assessment involves interrogating the product or output, evaluating it in relation to intelligent questions at a macro and micro level (Graesser, Pearson, & Magliano, 1995). Training in peer assessment seeks to develop this capability of asking intelligent, adaptive questions.

Peer assessment also involves increased time on task: thinking, comparing, contrasting, and communicating. Van Lehn, Chi, Baggett, and Murray (1995) suggested that peer assessment involves the assessor in reviewing, summarizing, clarifying, giving feedback, diagnosing misconceived knowledge, identifying missing knowledge, and considering deviations from the ideal. These are all cognitively demanding activities that could help to consolidate, reinforce, and deepen understanding in the assessor.

Chi (1996) distinguished among corrective feedback, reinforcing feedback, didactic explanations, and suggestive feedback. Giving simple correctional feedback (which only identifies an error and/or supplies the correct answer) challenges the assessor and the assessee minimally. Assessors should be trained to question, prompt, and scaffold rather than merely supply a notionally right answer.

For the assessed. When the criteria for assessment have been discussed, negotiated, used in practice, and clarified by all participants, greater clarity concerning what constitutes high-quality work is likely, which focuses assessee (and assessor) attention on crucial elements. Access to concrete examples of assessed work can also help students articulate the attributes of good and poor performance and promote the development of a vocabulary for thinking about and discussing quality. Peer assessment also involves *norm referencing*: enabling a student to locate himself or herself in relation to the performance of peers and to prescribed learning targets and deadlines. More accurate self-assessment might help avoid the adverse effects of overestimation or underestimation. Peer assessment might also reveal the next small step(s) needed to improve quality.

Peer assessment also makes available swifter feedback in greater quantity. In the event of misconception, it might prevent consolidation of confusion and the compounding of error upon error. Even where assessed products show no drastic misconceptions, peer feedback could prompt higher order or better quality thinking. Feedback is known to be associated with more effective learning in a range of settings. It yields higher rates of productive time on task and reduces cumulative error (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Crooks, 1988; Kulik & Kulik, 1988; Natriello, 1987). While peer feedback might not be of the high quality expected from a professional staff member, its greater immediacy, frequency, and volume compensate for this.

However, feedback is useful only when recipients act upon it. This has implications for the training of assessees. Simple summative, correctional, or didactic feedback is associated with much lower effect sizes than open-ended, suggestive, and formative feedback. Confirmatory or corroborative feedback is also important, since one might be correct without knowing whether or why one is correct. Different types of feedback can have different effects on different students. For example, there is some evidence that while instructor feedback is beneficial for students at low skill levels, it can be detrimental for students at high levels of skill (Teekell, 1989). There is also evidence that males and females respond differently to positive and negative feedback and differently to feedback from adults and peers (Dweck & Bush, 1976; Henry, 1979).

Overview. In a review of the wider literature on peer-assisted learning, Topping and Ehly (1998) noted that, cognitively, peer assessment might create effects by increasing a number of variables for assessors, assessees, or both. These variables 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, together 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 to the engineering of their closure through explaining,

Topping

simplifying, clarifying, summarizing, reorganizing, and cognitive restructuring. Feedback (corrective, confirmatory, or suggestive) could be more immediate, timely, and individualized. This might increase reflection and generalization to new situations, promoting self-assessment and greater metacognitive self-awareness. Cognitive and metacognitive benefits might accrue before, during, or after the peer assessment. Falchikov (1995a) noted that “sleeper” effects are possible.

Affect

Both assessors and assessees might experience initial anxiety about the process. However, peer assessment involves students directly in the learning process and may 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 interactivity, identification and bonding, self-confidence, and empathy for others.

Social and Transferable Skills

Peer assessment can develop teamwork skills and promote active rather than passive learning. It can also develop verbal communication skills, negotiation skills, and diplomacy (Riley, 1995). Learning how to give and accept criticism, justify one’s position, and reject suggestions are all forms of social and assertion skills. Student practice in peer evaluation could facilitate subsequent employee evaluation skills (Marcoulides & Simkin, 1991). Some projects specifically target peer assessment of transferable professional skills.

Systemic Benefits

Peer assessment can give students greater insight into institutional assessment processes (Fry, 1990). Students might thus develop more confidence in these processes and greater tolerance of the inevitable difficulties of discrimination at the margin. Alternatively, if institutional assessment procedures are inadequate, greater awareness of this among students could generate a positive press toward improvement. It has been contended that peer assessment is not costly in terms of teacher time and that peers are in ready supply (e.g., Fry, 1990). However, other authors (e.g., Falchikov, 1986) caution that there might be no saving of time in the short to medium term, since establishing good-quality peer assessment requires time for organization, 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 metacognitive benefits for staff as well as students. Peer assessment can lead staff members to scrutinize and clarify assessment objectives and purposes, criteria, and marking scales.

Disadvantages

Some authors have reported disadvantages or problems with their implementation of peer assessment (e.g., McDowell, 1995). Poor performers might not accept peer feedback as accurate. Students might not be willing to accept any responsibility for assessing their peers, especially initially, in a small socially cohesive group or if they see it as substitutional (Falchikov, 1995a). Byard (1989) noted that student groups can be inhibited and constrained, and the use and abuse of peer

power relationships should be monitored. Thus, peer assessment is not a universal panacea or necessarily a cheaper alternative to traditional assessment, although it might yield added value. Those new to peer assessment might also be concerned about issues of reliability and validity, to which we now turn. However, it should be noted that traditional assessment by tests or examinations, with multiple-choice and/or essay questions, is itself of doubtful reliability and validity, even for assessing surface learning of information (e.g., Newstead & Dennis, 1994).

Reliability and Validity of Peer Assessment

Thirty-one studies considered the reliability of peer assessment in higher education, most frequently through a focus on scores and grades awarded by peers rather than through a focus on more open-ended formative feedback. This was true even when the product to be assessed was a conceptually rich piece of original writing or an oral presentation as opposed to a simple test. The reason is doubtless that comparing quantitative indices is easy but raises other concerns. Furthermore, many purported studies of “reliability” appear actually to be studies of validity. That is, they compare peer assessments with assessments made by professionals rather than with those of other peers or the same peers over time. However, as Devenney (1989) pointed out, the role and function of teacher assessment might differ from that of peer assessment, so high reliability might not actually be necessary. Studies finding high reliability and low reliability are now considered separately in turn.

High Reliability

In a wide variety of subject areas and years of study, the products assessed have included essays (Catterall, 1995; Haaga, 1993; Marcoulides & Simkin, 1991, 1995; Orpen, 1982; Pond, Ulhaq, & Wade, 1995), hypermedia creations (Rushton, Ramsey, & Rada, 1993), oral presentations (Freeman, 1995; Hughes & Large, 1993a, 1993b), multiple-choice questions (Catterall, 1995), practical reports (Hughes, 1995), and professional skills (Korman & Stubblefield, 1971; Ramsey et al., 1996).

Of 25 studies comparing teacher and peer marks or grades, 18 (72%) reported acceptably high reliability, often expressed in correlation coefficients, percentage agreement, or measures of central tendency and variance, sometimes with indication of statistical significance (e.g., $r = .88$, Hughes & Large, 1993a, 1993b). A tendency for peer marks to cluster around the median was sometimes noted (e.g., Catterall, 1995; Taylor, 1995).

Structured assessment schedules were often used, sometimes with student involvement in their development (e.g., Pond, Ulhaq, & Wade, 1995; Stefani, 1992, 1994). Despite objective evidence of reliability, student acceptance (or belief in reliability) varied from high (Falchikov, 1995a; Fry, 1990; Haaga, 1993) to low (Rushton, Ramsey, & Rada, 1993). Haaga (1993) kept the process blind. Hughes and Large (1993a, 1993b) found that marks awarded by peers bore no relationship to the marks they received. Detailed formative feedback as well as the awarding of a grade was required by Falchikov (1995a).

Low Reliability

Lower or erratic reliability has been reported in the areas of essay writing (Mowl & Pain 1995), oral presentations (Taylor, 1995; Watson, 1989), peer-

mediated test taking (Hendrickson, Brady, & Algozzine 1987), and peer assessment of individual contributions to a group project (Mathews, 1994; Mockford, 1994). Again, student acceptance seemed unrelated to actual reliability (e.g., Hendrickson, Brady, & Algozzine, 1987). Mockford (1994) found good reliability for an overall peer mark but not for separate detailed components. Mowl and Pain (1995) found reliability unsatisfactory, despite training the participants, involving them in criteria generation, and supervising them carefully.

Self-Assessment Versus Peer Assessment

In self-assessment, Falchikov (1986) found younger students tended to be less reliable. More able students tended to undermark themselves, and average students tended to overmark themselves. 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 (cf. Hughes & Large, 1993a, 1993b). Shore, Shore, and Thornton (1992) found construct and predictive validity stronger for peer than for self-evaluations and 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 that 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 (.85) but lesser correlations between self-assessment and peer assessment (.61–.64). However, correlations between tutor assessment and self-assessment were even lower (.21), and those between tutor and peer assessment were modest (.34–.55). Self-assessment was associated with undermarking and clustering at the median.

Summary

Almost all of the 31 studies of the reliability and validity of peer assessment compared marks or grades, even for conceptually rich and various products. The majority of studies (18) suggest that peer assessment is of adequate reliability and validity in a wide variety of applications. However, a substantial minority (7) found the reliability and validity of peer assessment unacceptably low in particular projects. Peer assessment seems generally more reliable than self-assessment, but the two methods have some different sources of variance. Acceptability to students varies and is not a function of actual reliability. Of course, studies reporting low reliability might be intrinsically less likely to be published. There is a need for reliability and validity studies of detailed formative peer assessment (although Falchikov, 1995a, 1995b, involved this to some extent, and such a study is in at hand—Topping, Smith, Swanson, & Elliot, 1998).

The core of this review is subdivided according to the type of product that was peer assessed. Of the many options for subdivision, this seemed likely to be the most useful to the reader. The next six subsections of the review consider the literature on peer assessment through tests, marks, or grades; oral presentation skills; writing; group projects; professional skills; and computer-assisted peer assessment. Within the first three of these sections, descriptive studies are briefly mentioned initially before more detailed consideration of outcome studies.

Peer Assessment Through Tests, Marks, or Grades

Here studies are reviewed in which peers awarded marks or grades to their fellow students, whether for performance on simple multiple-choice tests or for performance on more complex activities and outputs such as oral presentations and written work.

Description

Descriptive studies include those of peer marking or grading of writing (Boud & Lublin, 1983), oral presentations (Conway, Kember, Sivan, & Wu, 1993; Earl, 1986; Falchikov, 1994), and group projects (Goldfinch, 1994; Goldfinch & Raeside, 1990; Rafiq & Fullerton, 1996). The wide variety of subject areas included engineering (Boud & Lublin, 1983; Rafiq & Fullerton, 1996), mathematics (Earl, 1986), optometry (Conway et al., 1993), social science (Falchikov, 1994), and art and design (Wright, 1995). Goldfinch and Raeside (1990), Goldfinch (1994), and Conway et al. (1993) devised a formula for allocating to an individual group member a percentage of the overall mark for the group's project. Other descriptive studies include Orpen (1982), Fry (1990), and Pond, Ulhaq, and Wade (1995).

Outcome

An early landmark study by Falchikov (1986) involved 48 biological science students in discussion and development of essay assessment criteria. They felt that the peer assessment process was difficult and challenging but that it helped develop critical thinking. A majority reported increased learning and better self-organization while noting that the process was time consuming. 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.

Peer assessment was applied to tests and to midterm and final exams by Ney (1989). This resulted in improved mastery of the subject matter and better classroom attendance. Watson (1989) found that after the introduction of peer assessment in one set of seminars, average marks were higher in the next set. Stefani (1994) had students define the marking schedule for peer-assessed experimental laboratory reports and reported learning gains from the overall process. The reciprocal group peer assessment researched by Mockford (1994) involved an overall evaluation followed by each group member assessing a specific aspect of the process design work of the other group in more detail. Participants reported learning benefits and greater clarity regarding assessment criteria.

A further application of her peer feedback marking scheme was conducted by Falchikov (1995a, 1995b; see the case study described earlier). 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 an 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 practicals increased in comparison to previous years, whose ability on entry was identical.

Summary

Peer assessment through tests, marks, or grades has been applied to many

Topping

different subject areas. Students find the process demanding but anxiety reducing. Learning gains in terms of test or skill performance or on subjective measures are frequently reported.

Peer Assessment of Oral Presentation Skills

Some studies in which marks or grades were awarded by peers for oral presentations were mentioned in the previous section (e.g., Falchikov, 1994, 1995a, 1995b; Pond, Ulhaq, & Wade, 1995; Watson, 1989).

Description

Peer assessment of undergraduate oral presentations has been described in mathematical modeling (Earl, 1986) and pharmacology (Hughes & Large, 1993a, 1993b). Wisker (1994) used the method in tutorless syndicate or peer groups, while Freeman (1995) compared assessments by groups of peers and faculty members.

Outcome

Heun (1969) compared the effect on student self-concept of peer and staff assessment of four public speeches given by students in a basic speech course. Relative to a control group, peer influence on the self-concept of students reached a significant level for the final speech, while instructor influence was nonsignificant across all four speeches. Mitchell and Bakewell (1995) found that 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. Assessment criteria were debated at length with the students. Participants felt that learning was enhanced and that the experience was relevant to peer appraisal skills in future work settings. Gains in trust and confidence in self and others were also identified, along with the development of a greater sense of responsibility.

Summary

In addition to improvement in marks and perceived learning (Falchikov, 1995a, 1995b; Watson, 1989), the relatively few outcome studies of peer assessment of presentations have found improved confidence and better presentation and appraisal skills. As yet, there are insufficient studies from which to draw any conclusions about the kinds or quality of feedback that might be most productive.

Peer Assessment of Writing

Descriptive

Although research in this area commenced early (e.g., Ford, 1973; Strickland, 1975) and is now voluminous, much of the relevant literature is descriptive. Peer assessment of writing, usually on a reciprocal basis, has been described in college composition seminars (Lewes, 1981; Orpen, 1982), undergraduate psychology (Camplse & Mayo, 1982; Haaga, 1993), technical writing classes (Samson, 1992), computing (Marcoulides & Simkin, 1995), and geography (Mowl & Pain, 1995; Pain & Mowl, 1996). Matching in pairs is usual, but Marcoulides and Simkin (1995) subjected each piece of writing to three independent peer assess-

ments. This could lead to more specific peer editing (Samson, 1992).

Lynch and Golen (1992) surveyed peer assessment practices in writing in business communication classes. Rothstein-Vandergriff and Gilson (1988) proposed a four-step model: (a) class-wide teacher-led discussion of a reading to model critical reading, (b) small-group discussion of the reading, (c) a collaborative writing assignment, and (d) individual writing assignments. Studies mentioned in previous sections (Boud & Lublin, 1983; Cavanagh & Styles, 1983; Falchikov, 1986; Hafernik, 1983; Lynch, 1982; Mangelsdorf, 1992; Marcoulides & Simkin, 1991; Pitts, 1988) also provided much useful descriptive organizational detail. Byard (1989) described and discussed the influence of student resistance and peer power relationships.

Peer assessment of writing has been used in English-as-a-second-language (ESL) contexts in several countries, especially in composition classes (e.g., Jacobs, 1989; Mendonca & Johnson, 1994; Obah, 1993; Witbeck, 1976). It has been noted that students use the peer assessor's feedback critically and selectively, an issue to address in training.

Outcome

In a business communication class, Roberts (1985) compared peer assessment with brief grammar reviews in groups of five and staff editing with no grammar reviews. Pretests and posttests showed a statistically significant difference in favor of the peer condition. The effects of teacher feedback, peer feedback, and self-assessment were compared by Birkeland (1986) among 76 technicians. No significant differences were found between conditions on test gains in paragraph writing ability. Similarly, Richer (1992) compared the effects of peer group discussion of essays with teacher discussion and feedback. Grading of 174 pretest and posttest essays from 87 first-year students indicated greater gains in writing proficiency in the peer feedback group ($p = .009$). Hughes (1995) compared teacher, peer, and self-assessments of written recording of pharmacology practicals, finding them equally effective.

Graner (1985) compared the effect of peer assessment and feedback in small groups with 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 practicing critical evaluation can have generalized effects on the evaluator's own work, even in the absence of any external feedback about the individual's own work. Perhaps time on task is a significant factor. Zhu (1994, 1995) investigated university freshman composition classes ($N = 169$) over one 15-week semester. Teacher-student interactive conference training was associated with more and more active and better quality peer feedback and with improved student attitudes toward peer revision and writing in general. However, the quality of writing done after peer revision did not differ significantly between groups.

Outcome studies of peer assessment of writing in ESL contexts are now considered together. Chaudron (1983) compared the effectiveness of teacher feedback, ESL peer feedback, and feedback from peers with English as their first language. Revised compositions were assessed for content, organization, vocabulary, language use, and mechanics. Students in all conditions showed a similar

Topping

pattern of improvement from first draft to revision. Working with 81 ESL college students 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. Devenney (1989) studied ESL students and teachers in Southeast Asia and suggested that the role and function of teacher evaluations differed from that of peer evaluations, with implications for the kind of comparisons reported earlier. Brock (1993) studied peer and computerized feedback on writing among ESL students in Hong Kong (see later discussion).

Summary

Peer assessment of writing is found in a wide range of subjects. Several studies have considered ESL applications. Peer assessment appears capable of yielding outcomes at least as good as teacher assessment and sometimes better. Formative feedback has been oral, written, and oral and written combined. There is some evidence that peer assessment without personal interactive feedback can be equally effective.

Peer Assessment of Group Work and Projects

Studies in which marks or grades were awarded by peers in the context of group work were mentioned in an earlier section (Goldfinch & Raeside, 1990; Mockford, 1994; Rafiq & Fullerton, 1996). Peer assessment has been used to help with the differentiation of individual contributions to small-group projects (Conway et al., 1993; Falchikov, 1993; Goldfinch, 1994; Mathews, 1994). For example, Keaten and Richardson (1993) worked with 110 speech communication students in 22 project groups. Individual contributions of group members were peer assessed via an instrument with six dimensions, although the relative emphasis given to quality of work was small. Student acceptability was high.

As early as 1981, Fineman had used peer assessment of group work with business administration undergraduates. Montgomery (1986) provided an interactionist process analysis of small-group peer assessment. Falchikov (1988) used self and peer group process assessment questionnaires to help promote competence, confidence, creativity, coping, and cooperation in a four-person group film-making project. In a related study of psychology students (Falchikov, 1993), group members and the lecturer negotiated self-assessment and peer assessment checklists of group process behaviors. Task-oriented behaviors proved easier to rate reliably than prosocial group maintenance behaviors.

Usher (1990) described peer assessment of communication skills in undergraduate group work in mathematical modeling. Johnson (1993) had advertising research students in project teams develop and use a peer assessment instrument, which enabled the instructor to compute numerical scores from qualitative evaluations. Abson (1994) had marketing research students working in self-selected tutorless groups use a simple 5-point rating scale on four criteria (cooperation, ideas, effort, and reliability). A case study of one group suggested that peer assessment might have made students work harder.

In engineering design, McKeown and Clarke (1995) integrated self-assessment and peer assessment in industrial product development team projects. Assessment

criteria were agreed on by the students and applied weekly. Total marks awarded to the project were then divided among individuals according to cumulated peer assessments. Higgitt (1996) used peer assessment with teams of students who were constructing interpretive field trials. Strachan and Wilcox (1996) used peer and self-assessment of group work to cope with increased enrollment in a third-year microclimatology course. Students found this fair, valuable, enjoyable, and helpful in developing transferable skills in research, collaboration, and communication. In summary, the outcome data on this type of peer assessment are limited to student perceptions.

Peer Assessment of Professional Skills

Applications

Peer assessment of professional skills can take place within the institution or on practical placements or internships. The latter case represents an interesting parallel to peer appraisal between staff in the workplace. It has been used by medical schools (Arnold, Willoughby, Calkins, Gammon, & Eberhart, 1981; Burnett & Cavaye, 1980; McAuley & Henderson, 1984), in preservice teacher training (Litwack, 1974; Reich, 1975), and in other professions. It has also been used in short practical laboratory sessions (e.g., Stefani, 1992). Application has been reported as well in more exotic areas, such as applied brass jury performances (Bergee, 1993) and a range of other musical performance arts (Hunter & Russ, 1995).

Medicine

In peer assessment of professional skills, acceptability to subjects is a major issue, as Jordan and Nasis (1992) found with nurses. 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 reported liking the exercise but feeling some anxiety about it. Ramsey and colleagues (1996) studied peer assessment of the professional performance of 187 medical interns. The process was acceptable to the subjects, and reliability was adequate despite the use of self-chosen raters.

Teaching

Franklin (1981) compared self, peer, and expert observational assessment of teaching sessions among preservice secondary science teachers. There were no differences between the groups in skill acquisition. A brief study by Turner (1981) yielded similar results. Yates (1982) used reciprocal paired peer feedback with 14 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 12 student teachers to give feedback to each other during 12 lessons in a 5-week practicum placement, but no training was given. Student

Topping

self-selection of partner proved no more likely to result in compatibility than random allocation. The participants reported the personal benefits to be improved self-confidence, praise and friendly support, confidentiality, mutual respect, and reduced stress. The benefits to their teaching included creative brainstorming and fine-tuning of lessons, resulting in improved organization, preparation, and delivery of lessons. Potential drawbacks included lack of trust and unbalanced, nonobjective, or dishonest feedback.

Counseling and Assertion Training

Peters (1978) studied the effects of video models, role play rehearsal, peer feedback, and remediation role play practice in the acquisition of counseling skills. All methods of training were equally effective in terms of written test and role play performance. In a highly controlled randomized study, Tillmann (1981) investigated peer feedback in assertiveness training with 204 female students. Significant treatment effects occurred on both behavioral and self-report measures. Greater amounts of peer feedback resulted in better assertive skills. During practice of helping skills in a class for 33 nonprofessionals, Teekell (1989) compared nonstructured peer feedback, peer feedback structured by a skills checklist, and structured feedback from the instructor. The last form of feedback had a highly beneficial effect for students at low skill levels but an equally detrimental effect on students at higher skill levels. Both peer conditions had minimal effects on all skill levels.

Business, Administration, and Commercial

Peer feedback in 114 insurance students was studied by Nilan (1983). The most important factors in terms of impact were source credibility, friendship influence, value of the feedback, and feedback style. Subjects preferred behaviorally defined performance dimensions with developmental suggestions to a vague rating scale. Calado (1994) studied business administration students to determine whether attitudes toward a peer feedback system were affected by satisfaction with the supervisor, attitude toward authority, fear of negative evaluation, self-esteem, and interpersonal trust. Some of these variables were statistically significant for some comparisons, but none were consistently so.

Summary

Peer assessment of professional skills shows adequate reliability (including in high-stakes areas such as medicine). However, outcome data are limited, often representing only participant perceptions. Nevertheless, peer assessment generally shows overall outcomes at least equivalent to teacher assessment. Observational schedules have some value in peer assessment of professional skills, while follow-up through self-monitoring merits further exploration.

Computer-Assisted Peer Assessment

Wider availability of word processing and electronic mail has created opportunities for formative peer assessment in electronic drafts prior to final submission, as well as distributed collaborative writing. For example, Downing and Brown (1997) described the collaborative creation of hypertexts by psychology

students, which were published in draft on the World Wide Web and peer reviewed via e-mail. Wider access to increasingly sophisticated speech-text software seems likely to affect peer assessment, especially of writing. In parallel, multiple-choice tests are increasingly administered and scored by computer, allowing detailed individualized feedback.

Rushton, Ramsey, and Rada (1993) and Rada, Acquah, Baker, and Ramsey (1993) reported on peer assessment in a collaborative hypermedia environment. The MUCH (Many Using and Creating Hypermedia) system had been used in nursing and computer science education. Good correspondence with staff assessment was evident, but the majority of computer science students were skeptical and preferred teacher-based assessment. Brock (1993) compared feedback from computerized text analysis programs 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 included significantly more words in postintervention essays.

Taylor (1995) had small groups of math students use computer tools for mathematical problem solving and then present their work to the class. Presentations were peer assessed in terms of clarity, originality, and presentational effectiveness. One paper (University of Portsmouth, 1995) described the creation of software to support peer assessment. It served organizational and record-keeping functions, randomly allocating students to peer assessors, allowing input by peer and staff assessors of marks given, integrating peer- and staff-assessed marks, calculating weighted final marks, and generating feedback for students. In short, various forms of computer-assisted peer assessment are now described in the literature, but few outcome data are yet available.

Quality Implementation of Peer Assessment

Organizational arrangements vary according to the type of peer assessment to be deployed, particularly the type of product. Good quality of organization is important for implementation integrity, in order to produce consistent and productive outcomes (Webb, 1995). Important general organizational factors in successful implementation emerging from the literature are summarized next, with exemplifying references.

Clarifying Expectations, Objectives, and Acceptability

Student expectations and objectives might be very different from staff objectives, and publicly explicit expectations and objectives might be different from those privately implicit. Expectations, objectives, and acceptability need to be clarified for all stakeholders and a collaborative and trusting ethos fostered. Although peer assessment can be highly acceptable to some students, there may be cultural differences, and acceptability might depend on confidentiality and the obviousness of formative intentions. To maximize acceptability, one should progress in steps that are absorbable and achievable by students, building confidence from experience and providing feedback at each stage (Calado, 1994; Cavanagh & Styles, 1983; Hafernik, 1983; Jacobs & Zhang, 1989; Williams, 1995; Zhang, 1995).

Matching Participants and Arranging Contact

How peer assessors and assessees should best be matched, and in what social constellation peer assessment should optimally occur, is discussed surprisingly little in the literature. Students might be matched with peer assessors whom they found credible or with whom they were already friends, or simply by random allocation. Most reports are of peer assessment in pairs, occasionally in small groups, and often reciprocal. Contact and discussion between assessor and assessee is typical but not essential (Boud & Lublin, 1983; Hafernik, 1983; Nilan, 1983).

Developing and Clarifying Assessment Criteria

Clarification and exemplification of the assessment criteria to be applied is seen as essential and student involvement in their development, elaboration, or simplification highly desirable. Inventories, checklists, response grids, and assessment criteria are often used, sometimes supported by more elaborate guides, model answers, and marking schedules. Within these elements, students prefer specific performance criteria to vague ratings (Boud & Lublin, 1983; Carlson & Roellich, 1983; Cavanagh & Styles, 1983; Christensen, Haugen, & Kean, 1982; Hafernik, 1983; Jordan & Nasir, 1992; Nilan, 1983; Pitts, 1988).

Providing Quality Training

Training for participants in actual application of the criteria is needed. Training might cover objectives, general organization, developing and using criteria and any associated materials, sustaining an effective group process, giving and receiving positive and negative feedback in different forms, action in response to feedback, and arrangements for evaluation. Video modeling has been found useful, and discussion opportunities are highly valued. However, direct experimental comparisons of trained and untrained groups are still awaited (Boud, 1995; Falchikov, 1993, 1995a; Jaques, 1984; Ogilvie & Haslet, 1985; Rothstein-Vandergriff & Gilson, 1988; Zhu, 1994, 1995).

Specifying Activities

Required or expected interactive behaviors should be clearly specified, explained, and preferably demonstrated with specimen products. These could include seeking specific error types, alternating specific tasks, identifying widely applicable metacognitive questions, and using developmental suggestions or prompts. Time limits might be set and different forms of feedback required (such as verbal and written). It should be clear whether the formative feedback is expected to affect assessees' next effort or whether they should rework their current effort. Articulating linkages to other ongoing teaching is important (Hafernik, 1983; Nilan, 1983; Witbeck, 1976).

Monitoring the Process and Coaching

The peer assessment activity should be monitored by staff while in process, especially when the participants are inexperienced. Further coaching or troubleshooting is likely to be necessary for at least some participants. Problems needing early detection include assessee errors or misconceptions unnoticed by the assessor, absent or faulty remediation, and the possibility of cheating and plagiarism.

Audio recording of students' discourse during peer assessment sessions has been used for monitoring. Process monitoring is, of course, very difficult when the participants do not actually meet to discuss (Boud & Lublin, 1983; Hafernik, 1983; Jacobs, 1989; Jacobs & Zhang, 1989; Zhu, 1994, 1995).

Moderating Reliability and Validity

Running checks on the reliability and validity of the peer assessments by staff need to be arranged, even if only on a sample, whether random or targeted. More than one peer assessment of a single output can lead to consideration of interassessor reliability. Alternatively, self-assessments can be contrasted with peer assessments (Boud & Lublin, 1983; Falchikov, 1993, 1995a).

Evaluating and Providing Feedback

Measures of improved student performance can be related to baseline rates of improvement prior to peer assessment or to the gains of comparison or control groups. More subjective measures to tap affective gains have included individual and group interviews and questionnaires. Self-assessment by assessors of the quality of their peer assessment has been used. Later follow-up to consider maintenance of gains is desirable, along with possible generalization of improved metacognition to other areas (Carlson & Roelich, 1983; Hafernik, 1983; Mangelsdorf, 1992; Pitts, 1988).

Overall Summary and Conclusions

Overall Summary

The literature on peer assessment between students in higher education is at an early stage of development, very variable in type and quality, and scattered and fragmentary in nature. Many reports are case studies from subject specialists. Studies of higher methodological quality are so varied in the type and organization of peer assessment investigated that a *best evidence synthesis* is still in the future. However, the current review does identify pertinent commonalities and differences to encourage fuller and more consistent reporting in the future and help promote more orderly, focused, coherent, and cost-effective onward research.

Peer assessment has occurred in a wide range of subject areas in relation to a wide range of outputs or products, including tests, writing, oral presentations, and skilled professional behaviors. Studies suggest that even simple quantitative feedback can have positive formative effects in terms of improved scores/grades and the subjective perceptions of participants. Studies have used both simple numerical and detailed open-ended peer assessment feedback; these are not mutually exclusive. Quantitative feedback seems more likely to be unidirectional, distant, and anonymous. Detailed feedback seems more likely to involve personal contact and to be reciprocal or mutual, personalized, and sometimes public.

Peer assessment seems equally likely to contribute to or not contribute to the assessee's final official grade. Most has been between students in the same year of study, and the relative ability of assessor or assessee has received little attention. Pair matching seems most frequent, but other constellations such as reciprocal group peer assessment are also reported. Peer assessment usually takes place in class during classtime (facilitating monitoring by staff) but can also occur at

Topping

least partially out of class, including in practical placements or internships outside the higher education institution. Studies often have failed to report whether involvement in peer assessment is compulsory or voluntary, although this might be expected to have an impact on acceptability and reliability. Awarding course credit or other incentives for participation was unusual.

Peer assessment seems adequately reliable and valid in a wide variety of applications, although virtually all of the current literature considers reliability of marks or grades rather than more detailed, formative assessment. Levels of acceptability to students are varied and do not seem to be a function of actual reliability. Students find peer assessment through tests, marks, or grades demanding but anxiety reducing. Learning gains in terms of test performance, skill performance, or subjective measures are frequently reported.

Peer assessment and feedback of a more detailed, open-ended nature have been associated with improved confidence and better presentation and appraisal skills. The relatively high number and quality of studies of peer assessment of writing suggest outcomes at least as good as teacher assessment, and sometimes better. Peer assessment of group and project work has been positive in the few studies to date, but evaluation has been largely in terms of student perceptions. Similarly, peer assessment of professional skills shows adequate reliability but limited outcome data, often in participant perceptions. However, these again show outcomes at least equivalent to teacher assessment. Various forms of computer-assisted peer assessment are emerging, although few outcome data are yet available.

Implications for Future Research

The next decade should bring a major expansion in the peer assessment literature. A more critical review, a best-evidence synthesis, and a meta-analysis should then become possible. Since peer assessment practices are so varied, future reports should include information on all 17 parameters in the typology and all 8 quality implementation factors, giving the basis for subsequent meta-analytic blocking. Also included should be information on participant characteristics and research design.

Participant characteristics. Considering both assessors and assessees, relevant variables might include familiarity and experience in peer assessment, geographical and/or cultural origin, chronological age, year of study, ability, and gender.

Research design. Relevant variables might include measurement of gains for assessors or assessees or both, the nature of measures and instrumentation ("hard" or "soft," single or multiple), degree of variance of peer assessments around the mean, the reliability and validity of the instrumentation (especially in relation to open-ended formative peer assessment), the incorporation of control groups that are demonstrably equivalent at pretest and/or the use of alternative treatment groups to control time on task, the extent to which maintenance of gains is planned for within the project via the use of self-monitoring or other means, and the degree to which longer term follow-up is pursued to check for washout of short-term gains or sleeper effects.

As the section on computer-aided peer assessment foreshadows, in the future nontraditional forms of assessment are likely to proliferate and expand, creating

new opportunities for peer assessment. For instance, portfolio assessment might become more widespread. Formative peer assessment of developing portfolios could form an extension of peer assessment of writing. Portfolios might be multimedia, incorporating photographs, other graphics, and videotaped and audiotaped material. A trend toward the electronic portfolio can be anticipated, with material in hypertext and hypermedia. It might prove easier to embed peer assessment in these new developments than to insert it into more traditional types of assessment where expectations can be concretized. There are signs (O'Donnell, 1998) that this is already happening in schools.

Implications for Future Practice

Practitioners wishing to establish peer assessment should consider the strengths and weaknesses of their personal and local context carefully. Consulting the typology can then inform consideration of the type of peer assessment that, in principle, best fits this context. For a first foray into peer assessment, practitioners might prefer methods already supported by a literature of reasonable quantity and quality (peer assessment through grades and peer assessment of writing). In any event, an action plan that addresses all of the factors in successful implementation described earlier should be formulated.

Conclusion

Peer assessment in higher education holds much promise and is becoming a mainstream idea (Brown & Dove, 1991). Organized, delivered, and monitored with care, it can yield gains in the cognitive, social, affective, transferable skill, and systemic domains that are at least as good as those from staff assessment. However, much further development and evaluation is needed, with improved methodological quality and fuller and more detailed reporting of studies. For practitioners, it is important that durable, cost-effective methods are identified with low innovation thresholds and the potential to be implemented on a large scale after piloting.

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* (Vol. 1, pp. 153–158). Southampton, England: 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, C. L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- 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.
- Birkeland, T. S. (1986). Variations of feedback related to teaching paragraph structure to technicians. *Dissertation Abstracts International*, 47, 4362.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher*

Topping

Education, 15, 101–111.

- Baud, D. (1995). How can peers be used in self assessment? In D. Boud (Ed.), *Enhancing learning through self assessment*. (Vol. 1, pp. 200–206). London & Philadelphia: Kogan Page.
- Boud, D., Churches, A., & Smith, E. (1986). Student self assessment in an engineering design course: An evaluation. *International Journal of Applied Engineering Education*, 2, 83–90.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18, 529–549.
- Boud, D., & Holmes, W. H. (1981). Self and peer marking in an undergraduate engineering course. *IEEE Transactions in Education*, E-24, 267–274.
- Boud, D., & Lublin, J. (1983). Student self-assessment: Educational benefits within existing resources. In G. Squires (Ed.), *Innovation through recession* (Vol. 1, pp. 93–99). Guildford, Surrey, England: Society for Research into Higher Education.
- 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, England: Standing Conference on Educational Development.
- 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, 273–287.
- Byard, V. (1989, March). *Power play: The use and abuse of power relationships in peer critiquing*. Paper presented at the Conference on College Composition and Communication, Seattle, WA.
- Calado, S. M. R. (1994). The impact of system design factors on users' reaction to peer feedback. *Dissertation Abstracts International*, 56, 1428.
- Camplese, D. A., & Mayo, J. A. (1982). How to improve the quality of student writing: The colleague swap. *Teaching of Psychology*, 9, 122–123.
- Carlson, D. M., & Roellich, C. (1983, April). *Teaching Writing Easily and Effectively to Get Results. Part 11: The Evaluation Process*. Paper presented at the Annual Meeting of the National Council of Teachers of English, Seattle, WA.
- 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* (Vol. 1, pp. 54–62). Coleraine, Northern Ireland: University of Ulster.
- Cavanagh, G., & Styles, K. (1983). Evaluating written work. *English Quarterly*, 16, 63–68.
- Chaudron, C. (1983, March). *Evaluating writing: Effects of feedback on revision*. Paper presented at the 17th Annual TESOL Convention, Toronto, Ontario, Canada.
- Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, 10, S33–S49.
- Christensen, L., Haugen, N. S., & Kean J. M. (1982). *A Guide to Teaching Self/Peer Editing*. Madison, WI: School of Education, University of Wisconsin-Madison.
- 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, 45–56.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.
- Devenney, R. (1989). How ESL teachers and peers evaluate and respond to student writing. *URELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, 20, 77–90.
- Downing, T., & Brown, I. (1997). Learning by cooperative publishing on the World

- Wide Web. *Active Learning*, 7, 14–16.
- Dweck, C. S., & Bush, E. S. (1976). Sex differences in learned helplessness: Differential debilitation with peer and adult evaluators. *Developmental Psychology*, 12, 1.
- Earl, S. E. (1986). Staff and peer assessment—Measuring an individual's contribution to group performance. *Assessment and Evaluation in Higher Education*, 11, 60–69.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education*, 11, 146–166.
- Falchikov, N. (1988). Self and peer assessment of a group project designed to promote the skills of capability. *Programmed Learning & Educational Technology*, 25, 327–339.
- Falchikov, N. (1993). Group process analysis—Self and peer assessment of working together in a group. *Educational & Training Technology International*, 30, 275–284.
- Falchikov, N. (1994). Learning from peer feedback marking: Student and teacher perspectives. In H. C. Foot, C. J. Howe, A. Anderson, A. K. Tolmie, & D. Warden (Eds.), *Group and interactive learning* (Vol. 1, pp. 411–416). Southampton, England: Computational Mechanics.
- Falchikov, N. (1995a). Peer feedback marking—Developing peer assessment. *Innovations in Education and Training International*, 32, 175–187.
- Falchikov, N. (1995b). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for learning in higher education* (Vol. 1, pp. 157–166). London: Kogan Page.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, 395–430.
- Fineman, S. (1981). Reflections on peer teaching and peer assessment: An undergraduate experience. *Assessment and Evaluation in Higher Education*, 6, 82–93.
- Ford, B. W. (1973). The effects of peer editing/grading on the grammar-usage and theme-composition ability of college freshmen. *Dissertation Abstracts International*, 33, 6687.
- 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.
- 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.
- 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.
- Gale, J. (1984). Self-assessment and self-remediation strategies. In E. S. Henderson & M. B. Nathenson (Eds.), *Independent learning in higher education* (pp. 99–140). Englewood Cliffs, NJ: Educational Technology.
- Goldfinch, J. (1994). Further developments in peer assessment of group projects. *Assessment and Evaluation in Higher Education*, 19, 29–35.
- Goldfinch, J., & Raeside, R. (1990). Development of a peer assessment technique for obtaining individual marks on a group project. *Assessment and Evaluation in Higher Education*, 15, 210–231.
- Graesser, A. C., Pearson, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9, 495–522.
- Graner, M. H. (1985). Revision techniques: Peer editing and the revision workshop. *Dissertation Abstracts International*, 47, 109.
- Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology*, 20, 28–32.
- Hafernik, J. J. (1983, April). *The how and why of peer editing in the ESL writing class*.

- Paper presented at the meeting of the California Association of Teachers of English to Speakers of Other Languages, Los Angeles, CA. (ERIC Reproduction Service No. ED 253064)
- 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, 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, 605.
- Heun, L. R. (1969). *Speech rating as self-evaluative behavior: Insight and the influence of others*. Unpublished doctoral dissertation, Southern Illinois University, Carbondale.
- Higgitt, D. L. (1996). The effectiveness of student-authored field trials as a means of enhancing geomorphological interpretation. *Journal of Geography in Higher Education*, 20, 35–44.
- Hughes, I. E. (1995). Peer assessment. *Capability*, 1, 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. (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* (Vol. 1, pp. 63–65). Coleraine, Northern Ireland: University of Ulster.
- Jacobs, G. (1989). Miscorrection in peer feedback in writing class. *Journal of Language Teaching and Research in Southeast Asia*, 20, 68.
- Jacobs, G., & Zhang, S. (1989, March). *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.
- Jaques, D. (1984). *Learning in Groups*. London: Croom Helm.
- Johnson, K. F. (1993). *Team peer evaluations: A student-generated quantitative measurement of group membership performance*. (ERIC Document Reproduction Service No. ED 361809)
- Jordan, J. L., & Nasid, D. B. (1992). Preferences for performance-appraisal based on method used, type of rater, and purpose of evaluation. *Psychological Reports*, 70, 963–969.
- Keaten, J. A., & Richardson, M. E. (1993). *A field investigation of peer assessment as part of the student group grading process*. (ERIC Document Reproduction Service No. ED 361753)
- Korman, M., & Stubblefield, R. L. (1971). Medical school evaluation and internship performance. *Journal of Medical Education*, 46, 670–673.
- Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58, 79–97.
- 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.
- 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* (Vol. 1, pp. 66–71). Coleraine, Northern Ireland: University of Ulster.
- Lewes, U. E. (1981). *Peer evaluation in a writing seminar*. (ERIC Document Reproduction Service No. ED 226355)
- 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.
- Lynch, D. (1982). Easing the process: A strategy for evaluating compositions. *College Composition and Communication*, 33, 310–314.
- Lynch, D. H., & Golen, S. (1992). Peer evaluation of writing in business communication classes. *Journal of Education for Business*, 5, 44–48.
- Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? *English Language Teaching Journal*, 46, 274–284.
- Marcoulides, G., & 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, 220–223.
- 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.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32, 302–313.
- McKeown, R., & Clarke, R. (1995). Product development team projects in final year engineering design. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing student learning through peer tutoring in higher education: Section 3—Implementing* (Vol. 1, pp. 79–86). Coleraine, Northern Ireland: University of Ulster.
- Mendonca, C. O., & Johnson, K. E. (1994). Peer-review negotiations/revision activities in ESL writing instruction. *TESOL Quarterly*, 28, 745–769.
- 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.
- Montgomery, B. M. (1986). An interactionist analysis of small-group peer assessment. *Small Group Behavior*, 17, 19–37.
- Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing—A case-study from geography. *Innovations in Education and Training International*, 32, 324–335.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155–175.
- Newstead, S., & Dennis, I. (1994). Examiners examined: The reliability of exam marking in psychology. *The Psychologist*, 5, 216–219.
- Ney, J. W. (1989). *Teacher-student cooperative learning in the freshman writing course*. (ERIC Document Reproduction Service No. ED 312659).
- Nilan, K. J. (1983). A study of recipient reactions to receiving developmental feedback from a peer group of co-trainees. *Dissertation Abstracts International*, 44, 29–30.
- Obah, T. Y. (1993). Learning from others in the ESL writing class. *English Quarterly*, 25, 8.
- O'Donnell, A. M. (1998). Peers assessing peers. In K. J. Topping & S. Ehly (Eds.), *Peer assisted learning*. Mahwah, NJ: Erlbaum.
- Ogilvie, J.R., & Haslett, B. (1985). Communicating peer feedback in a task group. *Human Communication Research*, 12, 79.
- 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 (Vol. 1, pp. 72–78). Coleraine, Northern Ireland: University of Ulster.
- Pain, R., & Mowl, G. (1996). Improving geography essay writing using innovative assessment. *Journal of Geography in Higher Education*, 20, 19–31.
- Patterson, E. (1996). The analysis and application of peer assessment in nurse education, like beauty, is in the eye of the beholder. *Nurse Education Today*, 16, 49–55.
- Peters, G. A. (1978). Effects of modeling, rehearsal, feedback, and remediation on acquisition of a counseling strategy. *Journal of Counseling Psychology*, 25, 231.
- Pitts, B. J. (1988). Peer evaluation is effective in writing course. *Journalism Educator*, 43, 84.
- Pond, K., Ulhaq, 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.
- Rafiq, Y., & Fullerton, H. (1996). Peer assessment of group projects in civil engineering. *Assessment and Evaluation in Higher Education*, 21, 69–81.
- Ramsey, P. G., Wenrich, M. D., Carline, J. D., Inui, T. S., Larson, E. B., & Logerfo, J. P. (1996). Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Academic Medicine*, 71, 364–370.
- 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.
- 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.
- Rothstein-Vandergriff, J., & Gilson, J. T. (1988, March). *Collaboration with basic writers in the composition classroom*. Paper presented at the annual meeting of the Conference on College Composition and Communication, St. Louis, MO.
- 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.
- Samson, D. C. (1992). Writing assignments for a graduate course in technical writing. *Journal of Technical Writing and Communication*, 22, 211–219.
- 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.
- Stefani, L. A. J. (1992). Comparison of collaborative self, peer and tutor assessment in a biochemistry practical. *Biochemical Education*, 20, 148–151.
- Stefani, L. A. J. (1994). Peer, self and tutor assessment—Relative reliabilities. *Studies in Higher Education*, 19, 69–75.
- 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, 343–353.
- Strickland, C. C. (1975). A peer-teaching technique for the two-year college composition student. *Teaching English in the Two-Year College*, 1, 71–75.
- 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* (Vol. 1, pp. 87-89). Coleraine, Northern Ireland: University of Ulster.
- Teekell, D. M. (1989). Effectiveness of peer feedback in helping skills training of nonprofessionals. *Masters Abstracts*, 28, 164.
- Tillmann, G. Y. (1981). The effects of peer feedback on assertive training outcomes with female adolescents. *Dissertation Abstracts International*, 41, 4278.
- Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.
- Topping, K. J., & Ehly, S. (Eds.). (1998). *Peer assisted learning*. Mahwah, NJ: Erlbaum.
- Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. J. (1998). *Formative assessment of academic writing between postgraduate students*. Manuscript submitted for publication.
- 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.
- University of Portsmouth. (1995). Transferable peer assessment. In National Council for Educational Technology (Ed.), *Using information technology for assessment, recording and reporting: Case study reports* (Vol. 1, pp. 73-78). Coventry, England: National Council for Educational Technology.
- Usher, J. R. (1990). Development of a staff and peer assessment scheme for groupwork in mathematical modelling. *Teaching Mathematics and Its Applications*, 9, 1-5.
- Van Lehn, K. A., Chi, M. T. H., Baggett, W., & Murray, R. C. (1995). *Progress report: Towards a theory of learning during tutoring*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: MIT Press.
- Watson, H. M. (1989). Report on the first year of research into developing an evaluative technique for assessing seminar work. *Collected Original Resources in Education (CORE)*, 13(2), Fiche 12 C1.
- Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis*, 17, 239-261.
- Williams, J. (1995). Using peer assessment to enhance professional capability. In M. Yorke (Ed.), *Assessing capability in degree and diploma programmes* (Vol. 1, pp. 59-67). Liverpool, England: Centre for Higher Education Development, Liverpool John Moores University.
- Wisker, G. (1994). Innovative assessment—Peer group and oral assessment. *Educational & Training Technology International*, 31, 104-114.
- Witbeck, M. C. (1976). Peer correction procedures for intermediate and advanced ESL composition lessons. *TESOL Quarterly*, 10, 321-326.
- 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* (Vol. 1, pp. 90-92). Coleraine, Northern Ireland: 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.
- Zhang, S. (1995). Re-examining the affective advantage of peer feedback in the ESL writing class. *Journal of Language Writing*, 4, 209.
- Zhu, W. (1994). Effects of training for peer revision in college freshman composition classes. *Dissertation Abstracts International*, 55, 951.
- Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication*, 12, 492-528.

Author

KEITH TOPPING is Director of the Centre for Paired Learning in the Department of Psychology at the University of Dundee, Dundee DD1 4HN Scotland. He develops and researches the effectiveness of methods for nonprofessionals (such as parents or peers) to tutor others in fundamental skills and higher order learning across a wide age and ability range and in many different contexts. He is also director of postgraduate professional training in educational psychology.

Received March 25, 1997

Revision received March 10, 1998

Accepted May 18, 1998