Peer Assessment : Learning by Judging and Discussing the Work of Other Learners

Keith J. Topping

University of Dundee, Dundee, Scotland, UK

Abstract

Peer assessment can be defined as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners, then learn further by giving elaborated feedback and discussing their judgements with peers to achieve a negotiated agreed outcome." It is organized in elementary (primary) and high (secondary) school classrooms and universities and colleges, but also in practitioner staffrooms and among teachers in training. There are many kinds of peer assessment, and factors in its variations are outlined. The evidence on the effectiveness of peer assessment is then reviewed. Four databases (ERIC, Science Direct, Scopus, and ZETOC) were searched using the terms "peer assessment" and school* or universit* or college. Some 230 papers of various quality were retrieved initially. For college/university, only reviews were reported here. Finally 43 papers of higher quality were selected as highly relevant. These mainly found that peer assessment was effective; only two finding otherwise. Some studies reported gains in meta-cognition and transferable skills in addition. Teachers in all sectors might consider or develop the use of the elaborated feedback form of peer assessment. Practical guidelines to aid this are given.

Corresponding author:
Keith J. Topping, University of Dundee, Dundee, Scotland.
E-mail: k.j.topping@dundee.ac.uk

Keywords

Peer assessment, discussion, elaborated feedback, elementary, high school, university, college

Introduction

What Is Peer Assessment?

This paper seeks to review the research literature on peer assessment in elementary and high schools and college/university in a way which is acceptable to academics but is also accessible to teacher practitioners.

Peer assessment came to widespread prominence about 20 years ago, and has become extremely popular in the last ten years. Associated terms are "peer feedback" or "peer review" or "peer response", but these do not mean quite the same thing. You will find peer assessment in the elementary (primary) classroom and in the high (secondary) school classroom. You will find it in universities and colleges. You will also find it in staff rooms, where teachers invite peer commentary on their own efforts from other teachers. And of course you will find it among teachers in training, who solicit peer feedback on their draft assignments.

There are simple and elaborated forms of peer assessment. The simple form is where teachers have peers grade or mark the work of one other peer, effectively using students as substitute teachers. This kind of use is a mainly a type of summative assessment, and teachers using it might be concerned about its reliability and validity, especially at the outset.

The elaborated form of peer assessment is much more developmental. Teachers have a

Topping K.J. *Interdisciplinary Education and Psychology.* 2017, 1:7.

1 of 17

peer independently assess the work of another peer, but the assessor then gives elaborated feedback and discusses their assessment with the assessee, giving a rationale and examples to support their judgements, and the assessee responds with their own thoughts. Subsequently (and after further independent reflection) the assessee revises the work in the light of the discussion. This kind of use is more clearly a type of formative assessment.

An early definition of peer assessment was "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners" (O'Donnell & Topping, 1998). Products to be assessed could include writing, oral presentations, portfolios, test performance, or other skilled behaviors.

However, this definition could include very simple peer assessment by peers assigning a score or grade to another student's work without ever giving any elaborated formative feedback. This elaborated feedback (and ensuing discussion and negotiation) is the interactive vehicle (with associated intellectual and social challenges) which leads to the occurrence of more learning. And this learning is for both members of the pair. So the new definition is more complex: "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners, then learn further by giving elaborated feedback and discussing their judgements with peers to achieve a negotiated agreed outcome".

Thus the formative view of peer assessment is emphasized here, in which the intent is to help students help each other plan their learning, identify their strengths and weaknesses, target areas for remedial action, and develop metacognitive and other personal and professional skills. Peer feedback is available in greater volume and with greater immediacy than teacher feedback. A peer assessor with less skill at assessment but more time in which to do it can produce an assessment of equal reliability and validity to that of a teacher.

Issues in Peer Assessment

This raises the question of who exactly is a "peer". Someone in the same class? Someone in the same year? Someone of the same ability in the subject in question irrespective of their chronological age? All of these are possible, except of course any member of staff is excluded and the "peers" all have the status of officially being learners. Equal status can be interpreted exactly (sometimes termed "near-peer") or with flexibility. In the latter case, a peer can be any student within a few years of schooling (far-peer).

Peer assessment usually deals with the products of learning – often writing, but also oral presentations, portfolios, drawings, and so on – but also with other associated skilled behaviors - learning behavior or wider social behavior - and sometimes encompasses both academic products and associated behavior. Thus peer assessment can not only consider the product of learning, but also the process behaviors which lead to learning.

Obviously, it is important that participating students are clear about the form and nature of peer assessment to be undertaken. Assessors may need to be given training in how to give positive, negative and neutral feedback, and maintain a balance between these. Equally, it is important that the recipient of feedback is ready to respond to it thoughtfully, deciding what points to accept and what not to accept, and using this selected information to either improve the existing piece of work (as in formative feedback) or future pieces of work (as in summative feedback). Often peer assessment is reciprocal, or at least most participants will be both assessed and assessor.

Initially peer feedback should highlight positive aspects of the work in question. Then it should move on to aspects that might be improved (one hesitates to say "negative"). For instance, this may involve not merely indicating the number of errors but saying exactly where they are, how they are wrong, and possibly suggesting how they might be put right. Then the pair can address features that may be errors but are open to discussion. Then the pair can discuss what aspects are missing from the product which perhaps should be there.

You might think that this form of elaborated peer assessment sounds rather time consuming, and you would be right. But during the process both assessor and assessee are presented with many intellectual challenges, set in the context of a piece of work that the assessee feels rather strongly about. As a result, both assessor and assessee have to think quite deeply, as well as negotiate a settlement, so not only are their thought processes stimulated, but their social skills as well. Consequently they both learn – and not just in one domain. So the

investment of time by a teacher can reap rich rewards.

Benefits and Problems of Peer Assessment

There are immediate benefits of peer assessment for learning and achievement (see the review of research literature below), but also longer term benefits with respect to transferable skills in communication and collaboration. There may also be ancillary benefits in terms of the self-regulation of the student's own learning. These benefits accrue to both assessors and assessees through the process of peer assessment. Peer assessment of learning and social behaviors sharpens and broadens the assessor's capabilities. Combining assessment of product and process can enhance student understanding of the consistency or mismatch between these, and different ways of learning beyond their own.

Of course, a problem with such peer assessment is that it may be less likely to be "correct" than teacher feedback. On the other hand, it is readily available in much larger quantity and more immediately than teacher feedback. Assessees just need to be somewhat cautious about peer feedback, and assessors cautioned that if they are uncertain they should not claim to be right, since they are not expected to always be right.

Once students are somewhat used to peer assessment, however, and have overcome their initial fears and hesitations, reliability is actually likely to be quite high (indeed, not that different from teacher reliability) (Topping, 2003a, 2009). Reliability can be increased by making model student work available to the assessors – a master version of correctness against which the work to be assessed can be compared. Additionally, assessors can be provided with scoring checklists or rubrics, to the design of which they have preferably contributed. Students need to know whether the peer assessments will be entered in any kind of high stakes assessment, such as end of year overall grades. Where this is the case, they will need to know what proportion of the total grade is determined by the teacher and what proportion by peer assessment.

Now let us consider a practical example so that the reader can get a feel of what peer assessment in action might look like. Bear in mind that while this is one practical example, it is only one example, and it should not condition your perception of what peer assessment is too precisely.

A Practical Example

A secondary school teacher in an English department wants to explore peer assessment of written analyses of a piece of Shakespeare. The hope is that this will engage the students more in what could otherwise become a mechanical exercise - writing only for the teacher. She is looking for more interactivity, better thinking, and greater generation of novel ideas. She discusses peer assessment with a colleague in her department, who is happy for her to take the initiative and will try it later in her own class if it works.

Knowing the students might balk at this new process, especially if it involves appearing unpleasant to their friends, the teacher takes care to introduce the idea gradually over a period of a few weeks. She divides the whole class into small groups to discuss the upcoming written task about Shakespeare. What might be the assessment criteria for this task? At first, the students are slow to respond, but eventually they warm up and generate a wide range of ideas about what a good essay would look like and what should be sought in the writing. The teacher works with them on synthesizing the criteria. Eventually, they have a reasonably short list of clear criteria. With one or two exceptions, the list is pretty much the same as the one the teacher herself would have used for assessing these papers, but the students do not know that.

The students are told that they will use these criteria to give feedback on each other's essays. They are divided into teams of three students of roughly similar ability in writing. This means that quite a few of the more able groups are all female and several of the less able groups are all male, but many in the middle are of mixed gender. The teacher takes care that each group contains no close friends (or enemies). Some students try to arrange to work with their friends, but the teacher tells them that would not be effective and is not part of the system.

The teacher then takes some time for training. She uses a similar, anonymous piece of writing done in the previous year. Displaying the piece together with the assessment criteria, she talks and shows by annotation how her mind is working as she makes assessment decisions about the manuscript, addressing each of the criteria. Then she invites the students to assess another previous year's paper in their groups, while discussing the task. They do

this quite quickly. As she circulates, she can see that some groups manage very well, yet others need some encouragement or coaching, which she gives. The teacher sets them to work on the new task. Each student completes their essay, knowing that it will be peer assessed by both of the other members of the group. This sharpens their motivation to make a good job of it. They also know that they will need to peer assess two papers themselves.

The groups convene and students complete their peer assessments without discussion in the group, referring to the written assessment criteria and a short list of useful tips about peer assessment. They know that they have forty minutes to complete the task of assessing two essays. As they work, the teacher goes round to look at what is being written and have a quiet word if anyone seems to be stuck. Most have finished within 30 minutes, and the teacher encourages them to make sure their assessments are all written down. She then lets them discuss the exercise until the end of the lesson, but not change their assessments. At the end of the lesson, each group hands in their written peer assessments to the teacher - six from each group of three people.

The teacher compares the two peer assessments for each essay. Where the two assessments are similar, she is content, and will pass both of these to the student who produced the work. Where they are very different, she notes that she will have to talk to this group next time they meet, and maybe even assess that essay herself before passing the assessments back to the student. Although there is sometimes a tendency for students' comments to verge towards the average (common with a first experience of peer assessment), she is aware that the next time the students will be more courageous. She is relieved that there is little sign of anyone giving a bad or good peer assessment just on the basis of personal preference.

At the next lesson, she gives the peer assessments back to the groups. An animated discussion ensues, not only about the nature of the comments on their analyses of Shakespeare but also about the merits and disadvantages of peer assessment. The students are a little dismayed that the teacher has not assessed the work herself but, on reflection, can see that their peer assessments are, by and large, just as useful in guiding revision. The class talks about how they might improve peer assessment the next time they do it. Somebody suggests that they apply it to another upcoming task, giving group presentations. The teacher agrees, noting that she has expended time on the training and peer assessment sessions but saved considerable time in not having to provide feedback on all these pieces of work herself overnight.

In terms of saving teacher's time (a not insignificant factor), you can see how while teacher's time is saved in assessing her/himself, more time is expended in discussing and agreeing the assessment criteria, training the assessors in giving feedback, coaching while peer assessment is occurring, and so forth. So in the short run teacher time is not saved, but re-allocated to other activities. In the medium to long term however, as the teacher and the class become more used to peer assessment, there may well be some saving of teacher time. However, there are likely to be benefits for teachers, as well as students. Peer assessment can lead teachers to scrutinize and clarify assessment objectives and purposes, criteria, and grading scales.

Now let us consider the different types of peer assessment (of which there are many), and be clear about the sources of variation.

Characteristics of Peer Assessment

There are many ways in which peer assessment can vary – and of course teachers are in charge of all these variables. However, they need to be clear about what they are.

As indicated above, a key difference is whether the peer assessment is formative or summative or both. Similarly, the peer assessment can be quantitative (assigning a number with respect to a grade) or qualitative (giving rich verbal feedback on positive and negative aspects and possibilities for improvement) or both. Thus, we see distinctions between formative and summative peer assessment, and between qualitative and quantitative peer assessment.Other differences between types of peer assessment are more subtle. For example, are the peer assessments on single pieces of work, or are they of several pieces of work? And are they on the same kind of product? The product or output assessed can vary - writing, portfolios, presentations, oral statements, and so on. Peer assessment can operate in different curriculum areas or subjects, which may impose different demands. For example, in

physical education classes, can peers be trained to investigate differences in the way the other student runs, or catches a ball, or throws a javelin, and so on? Will peer assessment be voluntary or compulsory? Will it be anonymous or not?

Clarification of the assessment criteria is essential, and peers may or may not be involved in establishing these criteria. Rubrics or structured formats for feedback may or may not be provided. Training in peer assessment may be given to assessors and/or assesses to a greater or lesser extent. Is feedback expected to be balanced between positive and negative, or only one of these? Is feedback expected to lead to opportunities to rework the product in the light of feedback, or is there no opportunity for this? Is feedback expected to include hints or suggestions for improvement? The nature of subsequent peer assessment activity may be very precisely specified or it may be left loose and open to student creativity. Does the interaction involve guiding prompts, sentence openers, cue cards or other scaffolding devices?

The participant constellation can vary, with consequent variation in joint responsibility for the assessed product. Assessors and the assessed may be individuals, pairs or groups. Directionality can also vary. Peer assessment can be one-way, reciprocal, or mutual within a group. Matching of students may be deliberate and selective or it may be random or accidental. It may take account only of academic factors, or also involve social differences. Assessors and assessed may come from the same year of study or from different years. They may be of the same ability, or deliberately of different ability. The amount of background experience students have in peer assessment can be very variable, and it may represent a considerable challenge to, and generate considerable resistance in, new initiates. If they have previous experience, it may have been positive, negative or both. Students from different cultural backgrounds may be very different in acceptance of peer assessment. Gender may make a difference, and thought should be given to the implications of same-sex or cross-sex matching. Of course, if there is no face-to-face contact (as in an online environment), gender may not be apparent, but this raises yet another source of variation.

Place can vary: most peer assessment is structured and occurs in class, but it can also be informal and occur outside of class. Similar variation occurs with respect to the time when the peer assessment takes place: How long are the sessions, how many sessions? The objectives for the exercise may vary - the teacher may target cognitive and/or metacognitive gains, or teacher time saving, or other goals. What degree of justification for opinions is expected of the assessor? Will all peer assessments be confidential to the assessing pair and the teacher, or will they be made publicly available?

The extent to which the process of peer assessment is monitored by supervisory staff (or whether the staff has no idea what actually occurred) is another question. The extent to which the reliability and validity of the peer assessment is moderated by supervising teachers is also an issue. Inspecting a sample of the assessments is particularly important where the assessment is summative. Is the task a simple surface task requiring limited cognitive engagement, or a highly complex task requiring considerable inference of the part of assesses, or does a simple initial task develop into increasingly complex tasks? In relation to this, what quantity and quality of feedback is expected, and is this elaborated and specific, or more concise and general? To what extent is it tending toward the objective and definitive, as it might be in response to a simple task, or to what extent more subjective, as it might be with a more complex task?

How are assessees expected to respond to feedback; are their revisions to be few or many, simple or complex? What extrinsic or intrinsic rewards are made available for participants? Is the peer assessment aligned with the traditional forms of assessment, or do all students have to sit formal examinations irrespective of their other skills? What transferable skills (e.g., social, communicative) might be measured as by-products of the process? Finally, is the peer assessment being evaluated, as one would hope with any new venture, or is its success or failure just assumed?

Thus, it is clear that peer assessment is not just one method. Labels can be given to some of these variations, distinguishing formative from summative peer assessment, qualitative from quantitative, structured from unstructured, unidirectional from reciprocal or mutual, same-year from cross-year, and same-ability from cross-ability peer assessment, for instance. These variations are summarized in Table 1 (below). Using this table, teachers will be able to

see what kind of peer assessment they intend to implement. Importantly, because all the variations are listed, teachers will not overlook any issue they should have considered. There are rather a large number of variables in the table, and some researchers have proposed clustering these. The difficulty is that different researchers propose different clusters, so I have left the list un-clustered.

Table 1: Variations in Peer Assessment

	Alternative A	Alternative B	Comment
1	Summative	Formative	or both
2	Quantitative grading	Qualitative feedback	or both
3	Single product	Several products	
4	Same kind of product	Different products	
5	Same curriculum area	Different areas	
6	Assessment criteria clear	Not clear	
7	Students involved	Student not involved	in defining criteria
8	Rubric used	Rubric not used	
9	Training given to peers	Not given	
10	Feedback positive	Feedback negative	or both
11	Feedback→improvement	No improvement	
12	Product reworked	Not reworked	
13	Scaffolding given	Not given	prompts, cues, etc.
14	Individuals	Pairs	or groups
15	One-way	Reciprocal	or mutual in group
16	Matching deliberate	Matching random	or matching accidental
17	Matching academic	Matching social	or both
18	Same year of study	Different year of study	
19	Same class	Different class	
20	Same ability	Different ability	in this subject area
21	Previous experience	No previous experience	
22	Experience positive	Experience negative	or both
23	Cultural expectations +ve	Negative	
24	Gender balance	Gender imbalance	ability, motivation, etc.?
25	Information technology	No IT	wholly or partly used?
26	In class	Out of class	or both
27	Length of sessions		
28	Number of sessions		
29	Objectives		Cognitive, metacognitive
30	Justification to peer	No justification	
31	Confidentiality	No confidentiality	to pair + teacher
32	Process monitored	Not monitored	
33	Reliability moderated	Not moderated	and validity
34	Task simple or complex		or simple→complex
35	Feedback expected		quantity + quality
36	Feedback objective	Feedback subjective	or both
37	Revisions many	Revisions few	
38	Intrinsic rewards	Extrinsic rewards	neither
39	Aligned	Non-aligned	with other assessment
40	Transferable skills	None measured	
41	Evaluated	Not evaluated	
42	Voluntary	Or Compulsory	
43	Anonymous	Non-anonymous	

Having clarified our concept of peer assessment and noted its diversity, we can now consider the evidence of whether (different kinds) of peer assessment actually work.

Evidence on Peer Assessment

The review below is initially based on existing reviews of the literature. However, much of the literature on peer assessment in concerned with students in university or college, and rather less with students in K-12 schools. This reflects the ease of access to their own students by university academics for research purposes. Schools can be elementary or secondary, and the organizational differences between these are such that implementation can be very different between the two environments. Consequently, I will take the two kinds of school separately, and discuss some individual papers as well as reviews.

Mostly the literature focuses on typical students, but a few papers focus upon students with special needs or various kinds of disability; these will be noted. For this review the literature was newly searched, with a particular focus on 1998 through 2017, although occasionally earlier classic items have been included. The literature before this has already been searched and the results reported in O'Donnell and Topping (1998). For the current chapter, four databases were searched: ERIC, Science Direct, Scopus, and ZETOC. The search terms were "peer assessment" AND school* OR universit* OR college. Some 230 papers were

retrieved initially. Qualitative, quantitative and mixed methods studies were included, but qualitative studies only where the methodology for data analysis was clearly explicated. Some retrieved items concerning irrelevant groups, and some lacked any kind of data. However, priority was given to studies which included an element of theorization as well as merely empirical data. For college/university, only reviews are reported here, and individual studies were discarded. This led to a final group of 11 reviews, 16 elementary studies and 16 high school studies.

Literature Reviews

An early review by Kane and Lawler (1978) considered research on three methods of peer assessment: peer nomination, peer rating, and peer ranking, noting that peer assessment could be reliable and valid. Peer nomination appeared to have the highest validity and reliability. Peer rating was the most useful of the three methods for feedback purposes.

A systematic literature review on the effects of peer assessment was reported by Van Gennip, Segers, and Tillema (2009). Fifteen studies since 1990 dealt with effects on achievement. However, only one of these studies included students from a school (a secondary school), the remainder consisting of university students. Peer assessment had positive effects. The authors developed four underlying constructs: psychological safety, value diversity, interdependence and trust. Psychological safety was defined as a belief that it is safe to take interpersonal risks in a group of people or the extent of confidence in the rest of the group. Value diversity refers to differences in opinion about what a team's task, goal or mission should be - it should be low for peer work to be effective. Interdependence has been long studied, but needs to be perceived by the participants rather than assumed by teaching staff. It requires that multiple perspectives are made explicit and students are individually responsible for an active contribution to group discussions. In respect of trust, several studies noted that students felt uncomfortable criticizing one another's work, or found it difficult to rate their peers, at least initially.

Van Zundert, Sluijsmans and van Merrienboer (2010) selected 27 papers, all of which studied university students. Peer assessment was improved by having trained and experienced peer assessors, and these also improved student attitudes towards peer assessment. Domain-specific skills were positively influenced by enabling students to revise their work on the basis of peer feedback, giving specific peer feedback formats, maintaining small group size, and allowing sufficient time for revision.

A review of research on the role of peer assessment in the elementary science classroom was undertaken by Hodgson (2010). The themes that emerged consistently were the need for a supportive classroom climate, the role of talk and discussion which was not all teacher-led, the importance of questioning in the process, and the richness of feedback.

Tillema, Leenknecht, and Segers (2011) considered what quality criteria were specifically relevant to peer assessment. One hundred and thirty-two studies of peer assessment were selected, together with 42 studies for a qualitative analysis. Nowhere was any distinction made between studies based in school, higher education or other settings. Studies were evaluated with regard to two quality criteria: 1) the recognition of educational measurement criteria, and 2) the consideration of student involvement in the assessment of learning. Where emphasis was placed on authenticity and future learning needs across the lifespan, peer assessment had much to recommend it in terms of generalizability, particularly utility in contexts beyond the present institution.

Hoogeveen and van Gelderen (2013) were the only authors to mainly consider school students, in analysing 26 studies of peer response on writing proficiency. They noted that several meta-studies had indicated that peer response was effective, but had not explored why. Many studies appeared to combine instruction in strategies, rules for interaction, and/or genre knowledge – and this seemed to be effective compared to individual writing.

Li, Xiong, Zang, Kornhaber, Lyu, Chung, et al. (2016) meta-analysed studies on peer assessment in digital platforms mainly in universities since 1999, finding an average correlation between peer and teacher ratings of 0.63 - moderately strong. This correlation was higher when: (a) the peer assessment was paper-based rather than computer-assisted; (b) the subject area was not medical/clinical; (c) the course was graduate level rather than undergraduate or K-12; (d) individual work instead of group work was assessed; (e) the

assessors and assessees were matched at random; (f) the peer assessment was voluntary instead of compulsory; (g) the peer assessment was non-anonymous; (h) peer raters provided both scores and qualitative comments instead of only scores; and (i) peer raters were involved in developing the rating criteria.

University and College

As mentioned above, many of the reviews primarily considered studies in higher education, so here we will add to this only by mentioning some reviews which were purely of higher education. Topping (1998) reviewed 67 studies and found that peer assessment was of adequate reliability and validity in a wide variety of applications. Peer assessment of writing and peer assessment using marks, grades and tests had shown positive formative effects on student achievement and attitudes. These effects were 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) were as yet more limited. Computer assisted peer assessment was an emerging growth area.

Sluijsmans, Dochy and Moerkerke (1998) conducted a review of self-, peer- and co-assessment, analysing 62 studies. They concluded that all of these were effective in developing professional competencies, and were often used in combination. Using these forms of assessment helped accelerate the development of a curriculum based on competencies rather than knowledge. The paper of Dochy, Segers and Sluijsmans (1999) was similar, concluding that the combination of new assessment forms encouraged students to become more responsible and reflective.

Falchikov and Goldfinch (2000) meta-analysed 48 quantitative peer assessment studies that compared peer and teacher marks to meta-analysis. Peer assessments were found to resemble teacher assessments more closely when global judgments based on well understood criteria were used rather than when marking involved assessing several individual dimensions.

Now let us look at individual papers of interest with respect to elementary and high schools.

Elementary (Primary) Schools

Surveys. Weaver (1995) surveyed 500 teachers regarding the writing process. These teachers generally found peer responses to be more effective than their own. In contrast, students stated they found the teacher's responses to be more helpful in all stages of writing, but they nevertheless improved when they received peer feedback about their writing.

Atkinson (2003) conducted two surveys about assessment across the curriculum in Scotland, where formative assessment had been used in primary classrooms for many years, with the same class of students and their parents. A mix of formal and informal assessment and self- and peer assessment were valued. Parents did not want just summative assessment.

Bryant and Carless (2010) conducted extensive interviews and classroom observations in a two-year case study of Hong Kong classrooms, which tended to be dominated by repetitive learning and many high stakes tests. Student perceptions of the usefulness of peer assessment varied according to the quality of peer feedback, peer language proficiency, and the novelty or repetitiveness of its processes. Teachers and students viewed peer assessment as having longer-term applications.

Effectiveness. Harrison, O'Hara, & McNamara (2015) noted that teacher assessment worked against the development of students into self-reliant people. Self- and peer-assessment were used with 523 students and their teachers. During self- and peer-assessment, students developed skills as critical, creative thinkers, effective communicators and collaborative team workers, becoming more personally productive and effective. Their self-awareness and self-reflection increased significantly. All of these were essential components of self-control and self-direction.

Boon (2016) studied children's uptake of feedback during peer assessment in primary school writing. Findings showed that pupils made better use of feedback if: it was task-involving and useful; there was sufficient time given for them to act on it and discuss it with their peers; and they were asked to reflect on how it has been used to improve the quality of the assessed work.

Yu and Wu (2016) sought to examine the individual and combined predictive effects of the

quality of online peer feedback provided and received on primary school students' quality of question-generation. Performance data from 213 fifth-grade students engaged in online question-generation and peer assessment for six weeks were analysed. Results indicated that the quality of peer feedback provided and received predicted students' quality of question-generation.

Efficacy of Different Modes of Presentation. Two studies examined the efficacy of different methods of applying peer assessment in elementary schools. Chin and Teou (2009) used concept cartoons with two parallel experimental classes of younger (9-10-year-old) and older (10-11-year-old) students. The cartoons presented opposing viewpoints about scientific ideas to stimulate talk and argumentation among students in small peer assessment groups. The dialogic talk and interactive argumentation of the students provided diagnostic feedback about students' misconceptions about scientific principles to the teacher, and was helpful in moving students towards better understanding.

In contrast to this verbal dialogic method, Yang, Ko and Chung (2005) developed a web-based interactive writing environment in a two-year longitudinal study. Peer assessment was used to provide constructive comments to foster students' ability to review and criticize other writers' essays, to enable students to review their own essay, and to encourage students to improve their writing skills. Students who participated in the writing environment, submitted many essays, interacted with other students online, and reviewed other essays, improved their writing skills.

Gender Effects. Yurdabakan (2011) conducted a study in a primary school fourth grade social sciences course with 46 participants (28 female and 18 male), their ages ranging from 9 to 10. Students scored their same and opposite sexes with respect to their contribution to group work and their learning levels. The compatibility between female student and teacher scores was higher than male student and teacher scores (the teacher was male).

Social Aspects. Studies using measures of academic achievement have found that students who score high on these are more accepted, less rejected and disliked by peers, viewed by teachers as less deviant, and engage in more positive interactions than those who score low on achievement (e.g. Malecki & Elliott, 2002). This may suggest that able students would make the best peer assessors. However, Bryan's (2005) research demonstrated that certain types of social skills interventions (particularly those focused on self-perceptions, self-attributions and locus of control) had consistent positive effects on academic achievement. The implication of this is that by engaging all students in peer assessment, it should be possible to raise the self-esteem and social-connectedness of rejected children and raise their academic achievement.

Frederickson and Furnham (2004) compared behavioral characteristics assessed by 867 typical classmates in mainstream middle schools (8-12-years-old) for children with moderate learning difficulties (MLD) (n=32) and socially rejected but not-MLD children (n=38), and their typical classmates (n=287). Systematic differences were identified between MLD and typical students in the peer-assessed behavioral characteristics, while there were no differences between typical and socially rejected students.

Four methods of assessing children's friendships were compared by Yugar and Shapiro (2001), addressing peer nominations (naming children you were friendly with), peer ratings (ranking children you were friendly with), and reciprocal peer nominations (naming children you were friendly with when they also named you). There was high agreement between reciprocal peer nominations and peer ratings, but regular peer nominations did not appear reliable. The extent to which social affiliation should be regarded as important when matching for peer assessment is an interesting one. Many practitioners do not match students who are already highly socially affiliated (since they may enjoy their relationship rather than using it to facilitate achievement), but also tend not match students who are strongly negatively affiliated.

Students with Special Needs. Peer assessment has been used successfully with special needs children (e.g., Scruggs & Mastropieri, 1998), with students as young as grade four (nine years old). Importantly, there are gains from functioning as either assessor or assessee. Studies in this category also have implications for matching students involved in peer assessment. O'Keefe (1991) examined sociometric surveys and peer assessments of social behavior in 51 third through sixth grade mainstream classrooms in order to identify

accepted and rejected intellectually challenged and non-challenged children. Rejected intellectually challenged children were perceived by their peers as aggressive/disruptive and/or sensitive/isolated. In contrast, accepted intellectually challenged children were perceived as sociable. The same relationships were found for typical children. This is encouraging, since it suggests that special needs status is no prohibition for involvement in peer assessment, although disruptive or isolated behavior might be.

Similarly, Rockhill and Asher (1992) examined gender differences in the types of behavior that distinguished between low-accepted children and their better-accepted classmates. Third- through fifth-graders (n=881) in five elementary schools completed a sociometric scale and a peer nomination measure. The same behaviors were important for boys and girls. Both aggressive and withdrawn low-accepted children received lower peer ratings for prosocial behavior. Children seemed to be blind to the condition or label of the problem individuals and took their behavior at face value. This is encouraging, though the effect may diminish as children grow older.

High (Secondary) School

High schools are more complex environments than elementary schools, and what works in the latter may not work in the former.

Perceived Value of Peer Assessment. Black and Harrison (2001) documented case studies of 12 teachers in six schools and reported some concern about whether peer assessment was practicable. Cultural differences were noted by Mok (2011), using interviews and classroom observation to report on moves towards peer assessment in the test-oriented special region of Hong Kong. Peer assessment has been recognized as enhancing student learning if sensitively implemented (Sebba, Crick, Yu, Lawson, Harlen, & Durant, 2008).

Accuracy of Peer Assessment. Tsivitanidou, Zacharia and Hovardas (2011) investigated secondary school students' peer assessment skills in Greece. Two classes of seventh graders (age 14, n=36, gender balanced) were anonymously assigned to reciprocally assess their peers' science web-portfolios. Interviews, video observation and peer assessors' feedback were used to find that the students had positive attitudes towards peer assessment and that they intended to implement it again.

In Taiwan, Sung, Chang, Chiou and Hou (2005) used progressively focused self- and peer assessment (PFSPA) procedures with students developing designs for new websites. Two classes of 76, 14–15-year-olds of above average ability assessed random websites produced by their colleagues. Self-assessment preceded peer assessment. The quality of the students' performance improved after the peer assessment activities. Similarly, Tseng and Tsai (2007) found that peer assessment scores were highly correlated with teacher scores.

Chang, Tseng, Chou and Chen (2011) who examined the reliability and validity of web-based portfolio peer assessment with 72 15-16-year-old students taking a computer course in a senior high school. Peer assessment scores were not consistent with teacher assessment scores and significant differences were found between peer assessment scores and end-of-course examination scores. The quality of training in peer assessment for these students was questionable and no assessment rubric was negotiated with the students.

Effectiveness of Peer Assessment. A Belgian team (Gielen, Peeters, Dochy, Onghena & Struyven, 2010) examined the effectiveness of certain characteristics of peer assessment feedback in a quasi-experimental, repeated measures study of 43, 13-year-old students. Written assignments showed that receiving justified comments in feedback was related to improvements in writing performance, but this effect diminished for students with higher pre-test performance. The effect of accuracy of feedback was less than the effect of justification.

The study by Sung et al. (2005) suggested that not only were students' peer assessments consistent with the assessments of teachers, but the quality of the students' work in developing new websites improved after the peer assessment activities as well. Similarly, Tseng and Tsai (2007) found that 184 tenth grade 16-year-old students significantly improved their projects for a computer course by engaging in successive rounds of peer assessment activities. The study also related the type of peer assessment feedback (reinforcing, didactic, corrective or suggestive) to subsequent performance in the student projects. Reinforcing peer feedback was useful in the development of better projects. Suggestive feedback was helpful in the beginning of peer assessment.However, didactic feedback, and to an extent corrective

Topping K.J. *Interdisciplinary Education and Psychology.* 2017, 1:7.

10 of 17

feedback, were negatively correlated with student achievement.

Chang and Tseng (2009) conducted a study in Taiwan on the use of peer assessment of web-based portfolios and its effect on student performance with 13–14-year-olds in two computer classes of 30 students each (one class intervention, the other control). There was no significant difference between groups in academic achievement and computer achievement. Similarly, Chang and Chou (2011) examined the effects of reflection quality in peer assessment on learning outcomes with 45 14-year-old students during a web-based portfolio process. The immediate influence of reflection quality on learning outcomes was small, but positive and statistically significant. Follow-up contrasts found reflection quality significantly related to an achievement test, work, and attitude outcomes.

Sung, Chang, Chang and Yu (2010) explored peer assessment in individual music performance with 116 seventh graders. Then 110 eighth graders had student-constructed web pages subject to peer assessments. Reliability and validity increased with the number of raters in both studies. Low- and high-achieving students tended to over- and underestimate the quality of their work, respectively. The discrepancy between the ratings of students and experts was higher in group-work assessments then in individual-work assessments.

Lu and Law (2011) studied 181 high school students engaged in online peer assessment. Peers graded and gave feedback, which was analyzed. Lu and Law found that the provision by student assessors of feedback that identified problems and gave suggestions was a significant predictor of the performance of the assessors themselves, and that positive affective feedback was related to the performance of assessees. However, peer grading behaviors were not a significant predictor of project performance.

van Zundert, Sluijsmans, Konings and van Merrienboer (2012) had 110 secondary school students study four integrated tasks, requiring them to learn a domain-specific skill and how to assess a fictitious peer performing the same skill. Additionally, the students performed two domain-specific test tasks and two peer assessment test tasks. Peer assessment skills were superposed on domain-specific skills and therefore suffered more when higher cognitive load was induced by increased task complexity.

Hovardas, Tsivitanidou and Zahcharias (2014) had 28 seventh graders anonymously assess each other's science web-portfolios. Peer assessors and an expert assessor used the same pre-specified assessment criteria. Peer assessees received feedback from two peers and the expert. Quantitative feedback differed between peer assessors and the expert and also between peer assessors - reliability was low. However, qualitative feedback was similar in that both included the structural components of written feedback. The majority of changes proposed by peer assessors were scientifically accurate.

Hsia, Huang and Hwang (2016) used a web-based peer-assessment method for performing arts activities. This was applied to 163 junior high students (experimental group and control group). Controls learned with a web-based streaming video-supported environment. Peer assessment using a rubric significantly improved the students' performance, self-efficacy and motivation. Peer assessment ratings were highly correlated with teachers' ratings on every performance item. Performance ratings were highly related to the students' self-efficacy. Students who learned with peer assessment were significantly more satisfied with the learning activity than controls.

Organizing Peer Assessment

This section attempts to provide teachers and consultants with a plan or template to address as they work on designing peer assessment projects. Obviously the characteristics of peer assessment outlined in Table 1 will also be a useful guide, and users should work through this table first. Good organization is perhaps the most important quality of implementation integrity, leading to consistent and productive outcomes. Of course, this is true of many other pedagogical interventions, and readers will find some of the points below have similarly wide relevance. Important planning issues evident in the literature (e.g., Topping, 2003a, 2009) are outlined below. This is intended as a summary, since we did not wish to overwhelm readers with too long a list of desiderata. Indeed, a summary of these summaries is given in Table 2 (below).

Table 2: Organizing Peer Assessment

	Things to Do
1	Work with colleagues
2	Clarify purpose, rationale, expectations, acceptability
3	Cognitive, attitudinal, social, or emotional gains?
4	Specify product(s) to be assessed
5	Introduce to students gently
6	Involve students in developing and clarifying assessment criteria
7	Match participants – same ability to start?
8	Arrange contact sessions – how many and how long?
9	Provide training, examples and practice
10	Give feedback and coaching during practice
11	Provide guidelines, checklists, other scaffolding
12	What for early and late finishers?
13	Monitor and coach
14	Examine the quality of peer feedback
15	Over time, moderate reliability and validity of feedback
16	Feedback this information to students
17	Evaluate

1. Seek to work with colleagues rather than developing the initiative alone.

2. Clarify purpose, rationale, expectations, and acceptability with all stakeholders. Are you aiming for cognitive, attitudinal, social, or emotional gains? Specify the nature of the products of learning to be assessed. Broach the idea with the students very early and, over time, seek their advice on and approval of the scheme.

3. Involve participants in developing and clarifying assessment criteria. Students need to be involved in developing the criteria for assessment in order to feel a sense of ownership and decrease any anxiety, even if they come up with something similar to what the teacher would have given them anyway. Small group discussion of teacher-proposed draft criteria should lead to a modest amount of suggested change.

4. Match participants and arrange contact. Generally aim for same-ability peer matching. If the peer partners are from the same class, roughly list them in order of ability in the subject of assessment, and pair the first two, the second two, and so on down the list (or the first three or four for peer response groups). Pairs or groups of students at the bottom of the list may be operating at the lowest level, but with some teacher support they may gain more than expected, as they will be involved in the same processes but at a simpler level.

5. Provide training, examples, and practice. Quality training will make a great deal of difference. Talk to students about what is expected of them, including the roles and behaviors expected of assessor and assessee. Then show them how to do it, perhaps by using a role play between two adults. Have the students practice peer assessment on a very short task selected for the purpose. While they practice, circulate to monitor their performance. Give feedback and coaching where needed.

6. Provide guidelines, checklists, or other tangible scaffolding. Some kind of written and/or pictorial reminders or clues to the process to be followed will help, e.g., a simple sheet with not more than eight reminders of what to do and how to do it.

7. Specify activities and timescale. Make clear what needs to be done, within what time-scale, and what records (if any) need to be kept. What of those who finish early – should extra peer assessment work be available or can they switch to some other kind of work? What of those who finish late - how can they be given timescales and reminders to keep them up to speed?

8. Monitor and coach. Whenever students are involved in peer assessment, keep a low profile and circulate among them, giving feedback and coaching as necessary.

9. Examine the quality of peer feedback. Particularly in the early days, check at least a portion of the peer assessments against your own assessments of the work. Choose a high, middle, and low ability student for this. Do not be surprised if the feedback is different from your own. The more feedback there is, the more chance it will be diverse. If it is very different, discuss this with the partners involved.

10. Moderate reliability and validity of feedback. Over time, keep consistent checks on the match between peer assessments (if more than one peer assesses the same piece of work), and on the relationship between peer and teacher assessments. Do not assume the teachers are any more reliable than the peers! You might want to match yours against the average of several peer assessments.

11. Evaluate and give feedback. Give the students information about your observations of their performance as peer assessors and your check on the reliability of their assessments. Unless they have this information, their ability to provide useful feedback will not change for the better.

Conclusion

Peer assessment has been shown to be effective in a variety of contexts and with students of a wide range of ages and abilities, particularly when organized and delivered carefully to meet the objectives specified (although of course this latter is true of other pedagogical techniques). The reliability and validity of peer assessments tends to be at least as high, and often higher, than teacher assessments (O'Donnell & Topping, 1998). Peer assessment requires training and practice, arguably on neutral products or performances before full implementation, which should feature monitoring and moderation. Given careful attention, a developmental process may be started that leads toward more sophisticated peer assessment, and the delivery of plentiful feedback that can help learners identify their strengths and weaknesses, target areas for remedial action, and develop metacognitive and other personal and professional skills.

Any group can suffer from negative social processes, such as social loafing (failing to participate), free rider effects (having the work of others accepted as one's own), diffusion of responsibility, and interaction disabilities (Salomon & Globerson, 1989). Social processes can influence and contaminate the reliability and validity of peer assessments. Peer assessments can be partly determined by friendship bonds, enmity, or other power processes, the popularity of individuals, perception of criticism as socially uncomfortable or even collusion to submit average scores, leading to lack of differentiation. Both assessors and assessees can experience initial anxiety about the peer assessment process.

Giving positive feedback first will reduce assessee anxiety and improve subsequent acceptance of negative feedback. In addition, students should be told that peer assessment involves students directly in learning, and should promote a sense of ownership, personal responsibility, and motivation. Teachers can also point out that peer assessment can increase variety and interest, activity and interactivity, identification and bonding, self-confidence, and empathy with others. Social factors require consideration by the teacher. When carefully organized, potentially negative social issues can be ameliorated and students can develop social and communication skills, negotiation and diplomacy, and teamwork skills. Learning how to give and accept criticism, justify one's own position, and reject suggestions are all useful, transferable social skills.

Reliability tends to be higher in advanced courses; lower for assessment of 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 (Karegianes, Pascarella, & Pflaum, 1980; MacArthur, Schwartz, & Graham, 1991). Peer assessments are generally more reliable when supported by training, checklists, exemplification, teacher assistance, and monitoring.

Peer assessment is unlikely ever to replace teacher or computer assessment as the main form of assessment. Quite apart from any other consideration, time would not permit the engagement of students in peer assessment for too large a proportion of their working week. However, I hope it has been made clear that peer assessment is about a lot more than assessment – it is also about improving the effectiveness of education generally and developing thinking skills (Topping, 2003b, 2009). It is also hoped that peer assessment is capable of engaging students much more effectively in self-regulation and developing other skills relevant to lifelong learning and work. Of course, such long-term developments have yet to be measured, and would be difficult to measure. Nonetheless, the logic of short-term measures and theoretical perspectives indicate that such a hope is not forlorn.

Many further details of the implementation of peer assessment in elementary and high schools will be found in Topping (2018), which includes numerous practical examples and a

discussion of how to evaluate peer assessment in these evidence-based times. Readers may also be interested in the accompanying volumes: Topping, Duran, & Van Keer (2017), Topping, Buchs, Duran & Van Keer (2017) and Duran & Topping (2017).

Funding Information

Conflict of Interest

The author has no conflict of interest to declare.

References

Atkinson, P. (2003). Assessment 5-14: What do pupils and parents think? Spotlight. *Education Resources Information Center document reproduction service,* ED480897.

Boon, S.I. (2016). Increasing the uptake of peer feedback in primary school writing: Findings from an action research enquiry. *Education* 3-13, 44(2), 212–225. http://dx.doi.org /10.1080/ 03004279. 2014.901984.

Black, P., & Harrison, C. (2001). Self- and peer-assessment and taking responsibility: The science student's role in formative assessment. *School Science Review*, 83(302), 43–49.

Bryan, T. (2005). Science-based advances in the social domain of learning disabilities. *Learning Disability Quarterly,* 28, 119–121.

Bryant, D.A., & Carless, D.R. (2010). Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation? *Educational Research for Policy and Practice,* 9(1), 3–15.

Chang, C.C., & Chou, P.N. (2011). Effects of reflection category and reflection quality on learning outcomes during web-based portfolio assessment process: A case study of high school students in computer application course. *Turkish Online Journal of Educational Technology*, 10(3), 101–114.

Chang, C.C., & Tseng, K.H. (2009). Use and performances of web-based portfolio assessment. *British Journal of Educational Technology*, 40(2), 358–370.

Chang, C.C., Tseng, K.H., Chou, P.N., & Chen, Y.H. (2011). Reliability and validity of web-based portfolio peer assessment: A case study for a senior high school's students taking computer course. *Computers and Education,* 57(1), 1306–1316. http://dx.doi.org/10.1016/j.compedu.2011.01.014

Chin, C., & Teou, L.Y. (2009). Using concept cartoons in formative assessment: Scaffolding students' argumentation. *International Journal of Science Education,* 31(10), 1307–1332. http://dx.doi. org/ 10.1080/09500690801953179

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), 331–350. http://dx.doi.org/10.1080 /03075079912331379935

Duran, D., & Topping, K.J. (2017). *Learning by teaching: Evidence-based strategies to enhance learning in the classroom.* London, UK: Routledge.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education. London,* UK: Routledge Falmer.

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

Frederickson, N.L., & Furnham, A.E. (2004). Peer–assessed behavioural characteristics and sociometric rejection: Differences between pupils who have moderate learning difficulties and their mainstream peers. *British Journal of Educational Psychology,* 74(3), 391–410.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction,* 20(4), 304–315. https://doi.org /10.1016 /j.learninstruc.2009.08.007

Harrison, K., O'Hara, J., & McNamara, G. (2015). Re-thinking assessment: Self- and peer-assessment as drivers of self-direction in learning. *Eurasian Journal of Educational Research*, 60, 75–88.

Hodgson, C. (2010). Assessment for learning in science: What works? *Primary Science,* 115, 14–16.

Hoogeveen, M., & van Gelderen, A. (2013). What works in writing with peer response? A review of intervention studies with children and adolescents. *Educational Psychology Review,* 25(4), 473–502. https://doi.org/10.1007/s10648–013–9229–z

Hovardas, T., Tsivitanidou, O.E., & Zahcharias, C.Z. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education,* 71, 133–152.

Hsia, L.H., Huang, I., & Hwang, G.J. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. *British Journal of Educational Technology,* 47(4), 618–632.

Kane, J.S., & Lawler, E.E. (1978). Methods of peer assessment. *Psychological Bulletin,* 85(3), 555–586.

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, 203–207.

Li, H., Xiong, Y., Zang, X., Kornhaber, M.L., Lyu, Y., Chung, K.S., et al. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245–264.

Lu, J., & Law, N. (2011). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 1–19. https://doi.org/10.1007/s11251–011–9177–2

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, 201–210.

Malecki, C.K., & Elliott, C.N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, 17, 1–23.

Mok, J. (2011). A case study of students' perceptions of peer assessment in Hong Kong. *ELT Journal*, 65(3), 230–239.

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

O'Keefe, P.F. (1991). Relationship between social status and peer assessment of social behavior among mentally retarded and nonretarded children. *Educational Resources Information Centre document reproduction service,* ED340500.

Rockhill, C.M., & Asher, S.R. (1992). Peer assessment of the behavioral characteristics of poorly accepted boys and girls. *Educational Resources Information Center document reproduction service*, ED346372.

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

Scruggs, T.E., & Mastropieri, M.A. (1998). *Tutoring and students with special needs*. In K. J. Topping & S. Ehly (Eds.), Peer-assisted learning. Mahwah, NJ: Lawrence Erlbaum.

Sebba, J., Crick, R.D., Yu, G., Lawson, H., Harlen, W., & Durant, K. (2008). *Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment*. Technical report. In: Research Evidence in Education Library. London, UK: EPPI-Centre.

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1998). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environments Research,* 1(3), 293–319.

Sung, Y.T., Chang, K.E., Chiou, S.K., & Hou, H.T. (2005). The design and application of a web-based, self-, and peer-assessment system. *Computers and Education,* 45(2), 187–202.

Sung, Y.T., Chang, K.E., Chang, T.H., & Yu, W.C. (2010). How many heads are better than one? The reliability and validity of teenagers' self and peer assessments. *Journal of Adolescence,* 33(1), 135–145.

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning - A review of research studies. *Studies in Educational Evaluation,* 37(1), 25–34.

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

Topping, K.J. (2003a). *Self and peer assessment in school and university: reliability, validity and utility*. In: M.S.R. Segers, , F.J.R.C. Dochy, & E.C. Cascallar (Eds.), Optimizing new modes of assessment: In search of qualities and standards. Dordrecht, NL: Kluwer Academic Publishers.

Topping, K.J. (2003b). Developing thinking skills with peers, parents & volunteers. *Thinking Classroom* (I.R.A.), 4(4), 27–36.

Topping, K.J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20–27 (themed issue on classroom assessment). Also in K. Cauley & G Panaozzo (Eds.), Annual Editions: Educational Psychology 11/12. New York, NY: McGraw-Hill.

Topping, K.J. (2018). *Using peer assessment to inspire reflection and discussion. Student assessment for educators series.* (Edn. J.H. MacMillan). London, UK: Routledge.

Topping, K., Duran, D., & Van Keer, H. (2016). *Using peer tutoring to improve reading skills: A practical guide for teachers*. London, UK: Routledge.

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation.* London, UK: Routledge.

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. http://dx.doi.org/10.1080/713611428

Tseng, S.C., & Tsai, C.C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers and Education,* 49(4), 1161–1174. https://doi.org /10.1016/j.compedu.2006.01.007

Tsivitanidou, O.E., Zacharia, Z.C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21(4), 506–519. https://doi.org /10.1016/j.learninstruc.2010.08.002

Van Gennip, N.A.E., Segers, M., & Tillema, H.M. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41–54. https://doi.org/10.1016/j.edurev.2008.11.002

Van Zundert, M., Sluijsmans, D., & van Merrienboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction,* 20(4), 270–279. https://doi. org/10.1016/j. learninstruc.2009.08.004

Van Zundert, M., Sluijsmans, D.M.A., Konings, K.D., & van Merrienboer, J.J.G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology*, 32(1), 127–145. http://dx.doi.org/10.1080/01443410.2011.626122

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

Yang, J.C., Ko, H.W., & Chung, I.L. (2005). Web-based interactive writing environment: Development and evaluation. *Educational Technology and Society,* 8(2), 214–229.

Yu, F.Y., & Wu, C.P. (2016). Predictive effects of the quality of online peer-feedback provided and received on primary school students' quality of question-generation. *Educational Technology & Society,* 19(3), 234–246.

Yugar, J.M., & Shapiro, E.S. (2001). Elementary children's school friendship: A comparison of peer assessment methodologies. *School Psychology Review*, 30(4), 68–85.

Yurdabakan, I. (2011). The investigation of peer assessment in primary school cooperative learning groups with respect to gender. *Education 3-13*, 39(2), 153–169. http://dx.doi.org/10. 1080/03004270903313608