

University of Groningen

Unfolding collaborative learning assessment literacy

Meijer, Hajo; Hoekstra, Rink; Brouwer, Jasperina; Strijbos, Jan-Willem

Published in:
Assessment & Evaluation in Higher Education

DOI:
[10.1080/02602938.2020.1729696](https://doi.org/10.1080/02602938.2020.1729696)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J-W. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education*, 45(8), 1222-1240. <https://doi.org/10.1080/02602938.2020.1729696>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Unfolding collaborative learning assessment literacy: a reflection on current assessment methods in higher education

Hajo Meijer, Rink Hoekstra, Jasperina Brouwer & Jan-Willem Strijbos

To cite this article: Hajo Meijer, Rink Hoekstra, Jasperina Brouwer & Jan-Willem Strijbos (2020) Unfolding collaborative learning assessment literacy: a reflection on current assessment methods in higher education, *Assessment & Evaluation in Higher Education*, 45:8, 1222-1240, DOI: [10.1080/02602938.2020.1729696](https://doi.org/10.1080/02602938.2020.1729696)

To link to this article: <https://doi.org/10.1080/02602938.2020.1729696>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 06 Mar 2020.



[Submit your article to this journal](#)



Article views: 2298



[View related articles](#)




[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

Unfolding collaborative learning assessment literacy: a reflection on current assessment methods in higher education

Hajo Meijer , Rink Hoekstra , Jasperina Brouwer  and Jan-Willem Strijbos 

Department of Educational Sciences, University of Groningen, Groningen, the Netherlands

ABSTRACT

Over the past two decades, curricula in higher education have increasingly incorporated collaborative learning. However, due to (a) large variations in students' domain-specific abilities (e.g. knowledge and/or skills) and the effort they invest into the collaboration and (b) teachers' limited knowledge about how to assess collaborative learning, two main challenges arise. The first challenge concerns ensuring construct validity of the assessment methods, that is, whether an assessment accurately measures students' domain-specific abilities. The second challenge originates from the potential of assessment methods to elicit student behaviour that is misaligned with the objectives of collaborative learning (e.g. free-riding behaviour). This paper aims to enhance teachers', researchers', and students' awareness for and need to develop what we refer to as 'collaborative learning assessment literacy'. In particular, we will discuss the two challenges in relation to three frequently used and discussed methods for assessing collaborative learning - group assessment, individual assessment, and group assessment combined with intra-group peer assessment - with specific attention to the purpose of assessment (i.e. formative and summative). Implications of the two challenges as well as their relation to other core components in the design of any collaborative learning setting (e.g. group constellation) will be discussed.

KEYWORDS

Collaborative learning; group work; assessment literacy; higher education

Over the past two decades, curricula in higher education have increasingly incorporated collaborative learning (Jin 2012; Flores et al. 2015). Although the literature contains a variety of terms and concepts to refer to students working and learning together - e.g. cooperative learning, collaborative learning, group work, peer learning, team learning, group learning activities, and group-based learning - we will use 'collaborative learning' as the umbrella term and Strijbos' (2016, 203) definition of collaborative learning as:

a learning phenomenon where individuals in a social constellation (e.g. group, team, or community) within a physical and/or virtual environment, interact on the same or different aspects of a shared task to accomplish implicit or explicit shared and individual learning goals (e.g. domain-specific knowledge or skills, social skills, etc.). Collaborative learning is structured by collaboration scaffolds (which can be faded if no longer needed) provided by an agent(s) within or outside of the social constellation (e.g. teacher, peer, self, technology) to guide interaction and increase the likelihood that social constellations and/or individuals can

CONTACT Hajo Meijer  hajo.meijer@rug.nl

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

accomplish their goals. An agent(s) within or outside of the social constellation diagnoses and/or evaluates the constellation's and/or individual's accomplishment(s) against criteria and standards.

Teachers may have didactic and pragmatic objectives for applying collaborative learning. Didactic objectives for using collaborative learning are, among others, to develop cognitive outcomes (e.g. knowledge), social outcomes (e.g. communication and collaboration skills), and motivational outcomes (e.g. attitudes) (Strijbos 2011). Pragmatic objectives for applying collaborative learning can be the sharing of learning materials (Van Aalst 2013) and reducing the time needed for teaching and grading students (Boud, Cohen, and Sampson 1999; Ahern 2007; Augar, Whitefield, and Winchester 2016). Although collaborative learning can lead to learning benefits compared to, for example, individual learning (Kyndt et al. 2013; Janssen 2014), students also experience problems with it (Pitt 2000; Payne et al. 2006; Carver and Stickley 2012). These problems appear related to the sub-optimal design of collaborative learning, such as minimal teacher guidance during collaborative learning and its assessment (Conway et al. 1993; Bolton 1999).

In our view two main principles are crucial to the design of collaborative learning: (1) adherence to individual accountability and positive interdependence as these are essential to any collaborative learning setting (Slavin 1980; Johnson 1981; Strijbos 2011), and (2) the alignment of eight core components for collaborative learning design as identified by de Hei, Strijbos et al. (2016). Regarding the first principle, individual accountability refers to the degree to which individual students are held accountable for their contribution to collaborative learning (Slavin 1980), whereas positive interdependence refers to the extent to which the performance of a single group member depends on the performance of other group members (Johnson 1981). Concerning the second principle, de Hei, Strijbos et al. (2016) developed the Group Learning Activities Instructional Design (GLAID) framework to guide teachers in designing collaborative learning settings to increase the likelihood that collaborative learning leads to the desired learning outcomes, by explicitly having teachers design the alignment between collaborative learning components. The framework consists of eight components: (1) interaction, (2) learning objectives and outcomes, (3) assessment, (4) task characteristics, (5) structuring, (6) guidance, (7) group constellation, and (8) facilities.

Out of the eight collaborative learning design components, the assessment component appears under-researched (Strijbos 2011, 2016; Forsell, Forslund Frykedal, and Hammar Chiriac 2020), as it was only explicitly mentioned in two of the 14 meta-studies on collaborative learning design that were analysed by de Hei, Strijbos et al. (2016). Assessment in education concerns the collection of evidence on the current state of student's domain-specific abilities (e.g. knowledge and/or skills) with two possible purposes: (a) summative, in which the assessment typically takes place at the end of the learning process and when the assessment outcome contributes to a final (high-stakes) assessment (i.e. evaluative), and (b) formative, in which assessment typically takes place during the learning phase to inform students about what needs to be done in the learning process without the assessment outcome contributing to a final (high-stakes) assessment (i.e. diagnostic) (Scriven 1967; Strijbos 2016).

As a part of the challenges teachers experience with the design of collaborative learning, they are also uncertain about and struggle with its assessment (Lejk, Wyvill, and Farrow 1997; Ross, Rolheiser, and Hogaboam-Gray, 1998; Ahern 2007; Gillies and Boyle 2010; Forslund Frykedal and Hammar Chiriac 2011; Hammar Chiriac and Granström 2012; de Hei, Strijbos et al. 2015; de Hei, Sjoer et al. 2016) as well as with the alignment of the assessment component with other collaborative learning design components (de Hei, Strijbos et al. 2016). This is evidenced, for example, by teachers struggling to align their objectives of collaborative learning (e.g. development of knowledge and/or skills) with their role in guiding and assessing collaborative learning (Ahern 2007; Forslund Frykedal and Hammar Chiriac 2011; de Hei, Strijbos et al. 2016; Opdecam and Everaert 2018).

It is hardly surprising that teachers experience the assessment of collaborative learning as challenging. Firstly, knowledge about how to assess collaborative learning is scarce in both

practical and research literature (Boud, Cohen, and Sampson 1999; Ahern 2007; Gillies 2007; Strijbos 2011, 2016; Brookhart 2013; Van Aalst 2013). Secondly, since students might differ in their domain-specific abilities (e.g. knowledge and/or skills) and the effort they invest into the collaboration, differences among students *within* and *between* groups may arise (Kagan 1995; Bacon, Stewart, and Silver 1999; Pitt 2000; Van Aalst 2013). On the one hand, variance among students - if not too large - is what makes collaborative learning worthwhile and contributes to its effectiveness since students can learn from each other's strengths and weaknesses (Bolton 1999; Roberts and McInnerney 2007). On the other hand, differences among students may pose two main challenges for any assessment of collaborative learning: (a) ensuring construct validity and (b) mitigating misaligned student behaviour.

Construct validity can be defined as a measure accurately measuring what it intends to measure, which Messick (1995) refers to as the content aspect of construct validity. Curricula in higher education typically emphasize the measurement of individual domain-specific abilities to decide whether students earn a degree (Boud, Cohen, and Sampson 1999; Nordberg 2006; Sharp 2006). Hence, the construct of interest when assessing students in higher education is their individual domain-specific ability (e.g. knowledge and/or skills). Thus, in the context of collaborative learning, construct validity entails whether an assessment method for collaborative learning accurately measures, and thus discriminates, between students' individual domain-specific abilities. With respect to misaligned student behaviour, assessment might elicit student behaviour that (a) counteracts the objectives of collaborative learning, e.g. individual assessment might decrease knowledge sharing and social support of group members; and/or (b) counteracts individual accountability and positive interdependence, e.g. increasing free-riding behaviour and/or devaluing the need to collaborate (Kagan 1995; Pitt 2000; Strijbos 2011, 2016).

Although it is sometimes argued that inconveniences with collaborative learning - often due to misaligned student behaviour - prepares students for possible unfair workplace practices (Alden 2011), we agree with Kagan (1995, 68) that "in the real world, there are many unfair practices - racial and age discrimination, unequal pay for equal work, and so on - but that doesn't justify unfair practices in the classroom". In addition, if teachers want to prepare students for unfair workplace practices and misaligned co-worker behaviour, it would make more sense to formulate learning objectives and outcomes related to how students can deal with unfair collaborative learning or workplace practices and subsequently design collaborative learning assessment in line with these objectives and outcomes, the task characteristics, and other design components of the GLAID framework (de Hei, Strijbos et al. 2016).

Notwithstanding that many teachers in higher education struggle with (a) ensuring construct validity and (b) mitigating misaligned student behaviour, they appear keen to improve their assessment literacy (Van den Bergh et al. 2006; Ahern 2007; De Hei et al. 2015). Assessment literacy can be defined as "the ability to design, select, interpret, and use assessment results appropriately for education decisions" (Quilter and Gallini 2000, 116), and in relation to collaborative learning it can be termed more specifically as 'collaborative learning assessment literacy'. Therefore, the current paper aims to provide more insight into the two collaborative learning assessments challenges (ensuring construct validity and mitigating misaligned student behaviour) to enhance teachers', researchers', and students' awareness for and need to develop collaborative learning assessment literacy in higher education. This will be done by discussing the two challenges for three frequently used (Lejk, Wyvill, and Farrow 1997), studied (Forsell, Forslund Frykedal, and Hammar Chiriac 2020) and discussed (Kagan 1995; Boud, Cohen, and Sampson 1999; Van Aalst 2013; Strijbos 2016) methods for assessing collaborative learning - (a) group assessment, (b) individual assessment, and (c) group assessment combined with intra-group peer assessment - with specific attention to the purpose of assessment (i.e. formative and summative).

In other words, we will focus on methods that assess the process(es) and product(s) of collaborative learning, without mixing in student performance on individual tasks. Although it is common practice to combine a score or grade for a group assignment and a score or grade for an

individual assignment(s) into a weighted grade for an entire course or unit (Lejk, Wyvill, and Farrow 1997; Forsell, Forslund Frykedal, and Hammar Chiriatic 2020), this practice has, in our view, implications for construct validity and (mis)alignment of student behaviour of a slightly different nature. That is, challenges with these practices are that (a) it assumes that individual tasks “validly compensate for a possible lack of individual effort and quality of contributions during the group tasks” (Strijbos 2016, 307) and (b) it is unclear how the grades should be weighted.

Group assessment of collaborative learning

Construct validity of group assessment

Group assessment is when all students in a group receive the same assessment by, for example, a group score, a group grade or comments at the group-level. Construct validity issues with group assessment of collaborative learning arise because work of individual students, which can be related to both the collaborative learning process and collaborative learning product, can become indistinguishable in the work of the group (Webb 1993; Kagan 1995; Nordberg 2008; Dijkstra et al. 2016; Forsell, Forslund Frykedal, and Hammar Chiriatic 2020). For example, the best predictor of a group score or grade appears to be the average ability of the group members (Bacon, Stewart, and Stewart-Belle 1998). Hence, while positive interdependence is fostered by group assessment, individual accountability is not. Given that curricula typically emphasize the measurement of individual domain-specific abilities to decide whether students earn a degree (Boud, Cohen, and Sampson 1999; Nordberg 2006; Sharp 2006), group assessment can be conceived of as problematic for the construct validity, since, for example, a group score or grade will appear as if it were an individual score or grade on students’ transcript of records. Likewise, it is a threat for the construct validity of a typical transcript of records (i.e. one that is centred around the individual) as it no longer solely represents students’ individual domain-specific abilities.

Because of the combination of within-group student differences and a lack of individual accountability, students with higher individual grades for individual assignments (e.g. written tests) might receive relatively lower group scores or grades for group assignments, whereas students scoring lower on individual assignments might receive relatively higher group scores or grades for group assignments (Lejk, Wyvill, and Farrow 1999; Almond 2009; Plastow, Spiliotopoulou, and Prior 2010; Moore and Hampton 2015; Harvey et al. 2019). In addition, it has been argued (Kagan 1995; Pitt 2000) and empirically shown (Bacon, Stewart, and Stewart-Belle 1998; Lejk, Wyvill, and Farrow 1999) that between-group differences might lead to situations in which students who are similar in their domain-specific ability might receive a different group assessment. That is, some students are placed within an ‘advantageous’ group with, for example, multiple highly motivated and/or high-achieving students, whereas others are placed within a ‘disadvantageous’ group with, for example, multiple poorly motivated and/or low-achieving students.

Empirical studies not only show that group assessment can be invalid, students can also complain about the construct validity of group scores or grades and can consider them as unfair (Conway et al. 1993; Willcoxson 2006; Ahern 2007; Carson and Glaser 2010). In addition, the more experienced students are with collaborative learning, the less they claim that all students in a group deserve the same score or grade (Barfield 2003). Teachers may also struggle with the construct validity of group assessment and therefore indicate, for example, to limit the contribution of the group assessment in making up the total grade of a course (Ahern 2007; Augar, Whitefield, and Winchester 2016). Invalid formative group assessments may result in feedback that does not correspond to each group member’s individual learning process and therefore could be suboptimal in terms of student learning benefits. Invalid summative group assessment might result in low-achieving students passing a course while individually not meeting the requirements and standards (Felder and Brent 2001). Conversely, high-achieving students might not pass a course or pass with a lower than the desired grade and/or a grade that does not

represent their individual domain-specific abilities, due to their dependence on fellow students' comparatively lower domain-specific abilities (Nordberg 2006).

Behavioural (mis)alignment of group assessment

Group assessment aligns with the objectives of collaborative learning (e.g. letting students work together), on the hand, because it fosters positive interdependence. That is, students depend on each other in order to succeed in the collaborative assignment, which can lead to, for example, students helping each other. On the other hand, due to the lack of individual accountability, group assessment can elicit student behaviour that is misaligned with the objectives of collaborative learning. Firstly, it might elicit free-riding (when one or more students within a group do not perform a fair share of the work, either because a student is not motivated or assumes that the group is better off without his or her contributions), social loafing (when one or more students invest less effort compared to working individually), and the sucker effect (when students are taken for a free-ride by other group members who invest less effort resulting in one student doing all the work, thus being a 'sucker') (Salomon and Globerson 1989).

Secondly, especially when the stakes are high, students might adopt a performance-oriented approach rather than a learning-oriented approach (Kagan 1995; Pitt 2000). That is, to maximize a group assessment (e.g. score or grade), the most optimal strategy to achieve the best possible group performance might be to let every group member do what he or she does best and thus divide the group task into individual tasks, combining them in the end (Pitt 2000; Macdonald 2003; Knight 2004; Bacon 2005). Not only might this behaviour be misaligned with the objectives of collaborative learning, it might also lead to less learning (Bacon 2005), which in turn is also misaligned with the objective of collaborative learning. Moreover, when the within-group student differences are large in terms of domain-specific abilities, with, for example, high achievers who strive for the best possible assessment (e.g. high group grade) and low achievers who are just interested in passing the course regardless of the assessment outcome, high achievers might make additional efforts in order to ensure the best possible assessment (Conway et al. 1993; Kagan 1995; Pitt 2000). Additional efforts could, for example, be that the high achievers take over tasks from the low-achievers or that they supervise the low-achievers by giving feedback and/or correcting the work. From a behavioural alignment perspective, a high-achiever supervising a low-achiever represents positive behaviour as it resembles peer tutoring by which students collaborate, help each other and thereby learn from each other. From a construct validity perspective, however, this behaviour might be a threat to the construct validity of group assessment as it results in an unequal contribution to and responsibility for the group assignment which can lead to an assessment not reflecting students' individual domain-specific abilities.

Hence, group assessment might lead to student behaviour that is misaligned with the objectives of collaborative learning (Kagan 1995; Pitt 2000). That is, group assessment might elicit behaviour such as free-riding and adopting a performance-oriented approach, which in turn might lead to, for example, unequal division of tasks, unequal contributions, and unequal responsibilities for the group assignment (Pitt 2000; Macdonald 2003; Knight 2004; Bacon 2005). Although this might be the case for both formative and summative assessment situations, the likelihood of adopting a performance-oriented approach is expected to be higher for summative assessment due to its high stakes nature.

Individual assessment of collaborative learning

Construct validity of individual assessment

Individual assessment of collaborative learning is when students in a group receive a personalized assessment (e.g. individual score, grade, or comments at the individual-level) for their

individual responsibilities and/or (sub)tasks within the collaborative learning assignment. Hence, in contrast to group assessment, individual assessment of collaborative learning fosters individual accountability, whereas positive interdependence is not fostered. Therefore, given that curricula typically emphasize the measurement of individual domain-specific abilities (Boud, Cohen, and Sampson 1999; Nordberg 2006; Sharp 2006), a higher construct validity - compared to group assessment - might be assumed.

On the one hand, the individual assessment that students receive of their contribution(s) to a collaborative learning assignment might indeed better reflect their individual domain-specific abilities and thus better discriminate among students within a group. On the other hand, the individual assessment is most likely not entirely detached from the collaborative setting because students' individual performance is likely affected by (a) the collaborative setting and behaviour of the fellow group members irrespective of the intensity of the collaboration, (b) a potential benefit from exchange of knowledge, and (c) the development of social skills from collaborating cannot be readily discarded (Strijbos 2011). Hence, the seemingly individual formative or summative assessment of collaborative learning is actually 'purposely confounded' by the design and objectives of collaborative learning. In other words, what is assumed to be an individual assessment of collaborative learning might at best be qualified as a 'collaboration-moderated individual assessment of collaborative learning'.

However, since students ultimately have to perform individually, individual assessment of collaborative learning is, compared to group assessment, most likely to be more valid, making it less susceptible to low achievers benefiting from high achievers or high achievers suffering from low achievers (Lejk, Wyvill, and Farrow 1999; Almond 2009; Harvey et al. 2019). Consequently, formative individual assessment of collaborative learning might lead to an individual score, grade or comments at an individual level that likely better aligns with each group members' individual learning process and which thereby might increase their individual learning. Summative individual assessment of collaborative learning is less likely to result in students passing a course who individually do not meet the required abilities (Felder and Brent 2001) and to high-achieving students not passing a course or passing with a lower than the desired grade/and or grade that does not represent their individual domain-specific abilities (Nordberg 2006).

Behavioural (mis)alignment of individual assessment

Individual formative and summative assessment of collaborative learning fosters individual accountability, which counteracts group assessments' problems with free-riding, social loafing, and the sucker-effect. However, as a result of the lack of positive interdependence with individual assessment of collaborative learning, both formative and summative individual assessment of collaborative learning do not function as an incentive for students to engage in genuine collaboration. Even more than with group assessment, all that it takes to succeed in the case of individual assessment of collaborative learning is that students arrange superficial and high-level organisational matters such as dividing subtasks and deciding who combines the separate pieces together (Macdonald 2003; Knight 2004; Bacon 2005). That is, students might work individually on their subtasks without any genuine integrative group task-related collaboration with their fellow group members.

It has been shown that individual assessment of collaborative learning can create rivalry among group members which might hamper collaboration within groups (e.g. impaired information sharing) (Hayek et al. 2015). As a consequence, irrespective of whether formative or summative individual assessment of collaborative learning is applied, it is questionable whether students' behaviour aligns with the behaviour that the teacher aimed for; and, thus, whether the objectives of collaborative learning are achieved (especially when collaborative learning aims for the development of skills such as communication and social skills). In addition, as with group assessment, individual assessment of collaborative learning might lead to less learning since students might only focus on one or several but not every aspect of the group assignment (Bacon

2005). Finally, if the content of the collaborative assignment hardly fosters collaboration, individual assessment of collaborative learning is likely to turn the collaborative assignment into an individual assignment in disguise.

Group assessment combined with intra-group peer assessment of collaborative learning

Construct validity of group assessment combined with intra-group peer assessment

Peer assessment of collaborative learning can occur in two major formats: intra-group peer assessment and inter-group peer assessment (Sivan 2000). In intra-group peer assessment each group member assesses all fellow group members (and sometimes also themselves), whereas in inter-group peer assessment a group or individual members of that group assesses the product of another group. Of the two, intra-group peer assessment is the most common, as it (a) aligns quite naturally with collaborative learning and (b) can provide insight into the collaborative process which is typically not accessible to the teacher (Onyia 2014; Strijbos 2016). Hence, intra-group peer assessment is often combined with a group assessment (e.g. group score) for the group product. The group product can then be moderated into an individual score or grade for each group member via the peer assessment scores by fellow group members regarding, for example, each group member's contribution to the group task, their specific responsibilities, and/or the effort they invested. For example, a student who contributed less than fellow group members, which is hard to identify for teachers, can be detected and their group score or grade 'corrected' via intra-group peer assessment (Cheng and Warren 2000). Therefore, group assessment combined with intra-group peer assessment could be a solution for group assessments' limitations in light of individual accountability (e.g. resulting in free-riding behaviour) and individual assessments' limitations in light of positive interdependence (e.g. resulting in limited or no incentive to engage in genuine collaboration).

Intra-group peer assessment can stimulate participation of students in groups, and reduce free-riding behaviour and social loafing (Johnston and Miles 2004; Shiu et al. 2012; Sridharan, Muttakin, and Mihret 2018). As such, intra-group peer assessment might increase the construct validity of a group assessment. However, challenges for ensuring construct validity arise as well when combining group assessment with intra-group peer assessment. Firstly, intra-group peer assessment typically focuses on the collaborative process rather than the collaborative product (Forsell, Forslund Frykedal, and Hammar Chiriac 2020). Similar to the assessment of an individual's learning product not necessarily informing teachers about students' individual learning process (Gibbs 1995), the assessment of the collaborative process does not necessarily inform teachers about the collaborative product. That is, although students might work well together and every group member does his or her fair share - resulting in similar intra-group peer assessment scores for their contribution to the collaborative process - some students might deserve a higher score or grade based on the quality of their contribution to the group product. The group product, however, is typically still only scored or graded with a group assessment, despite the criteria on the quality of the contribution to the collaborative product also being part of the intra-group peer assessment (Cheng and Warren 2000; Lejk and Wyvill 2001; Johnston and Miles 2004). Yet, in that case, similar challenges might arise as with group assessment: students with higher scores or grades for individual assignments (e.g. written tests) might receive relatively lower scores or grades for group assignments, whereas lower scoring students might receive relatively higher scores or grades for group assignments (Lejk, Wyvill, and Farrow 1999; Almond 2009; Harvey et al. 2019).

Secondly, students' ability to validly assess their peers can be compromised. For example, intra-group peer assessments can be influenced by friendships, peer pressure, harsh marking as a consequence of conflicts within a group or students disliking each other, and deliberately

down or up-grading fellow group members to enhance their own relative profit from peer assessment (Bushell 2006; Willcoxson 2006; Carver and Stickley 2012; Jin 2012; Ohaja, Dunlea, and Muldoon 2013). Moreover, Diprose, Judd, and Morris (1997) found that when students assessed their peers' contribution as poorly, they feared that their own contribution would be assessed as poorly in retribution. These biases are known as the 'reciprocity effect', which states that the relationship between peers might influence their assessment of their peers (Magin 2001). Because of the aforementioned biases, it is no surprise that students can perceive intra-group peer-assessment as unfair (Roskams 1999; Kaufman and Schunn 2011; Jin 2012). However, Sridharan, Tai, and Boud (2019) found that students are able to assess peers validly in a formative context, but not in a summative context. Hence, it seems that students are reluctant to assess peers honestly when the stakes are high because they do not feel comfortable to penalise their peers and fellow group members (Ohaja, Dunlea, and Muldoon 2013; Hastie 2018).

Thirdly, the formula by which the moderation commonly is computed is still subject to debate. In general, the formula used is typically based on the PA-factor by Goldfinch (1994); occasional minor variations aside (Strijbos, Stegmann, and Sluijsmans, 2017). Although this formula is used, for example, as part of the SPARK_{PLUS} system for summative purposes (see Wu, Chanda, and Willison 2014), it is still subject to the aforementioned biases; hence, several corrections have been proposed for subjectivity and dishonesty of peer assessment scores (Li 2001), assessing oneself higher or lower than fellow group members (Bushell 2006), and large variations in individual scores due to the formula used for the moderation (Sharp 2006; Neus 2011; Nepal 2012; Spatar et al. 2015). However, to date there are no agreed upon practices or formulas to convert a group score or grade into individual scores or grades with the help of intra-group peer assessment.

In sum, although group assessment combined with intra-group peer assessment seems like a solution to group and individual assessments' limitations, several challenges arise: (a) the collaborative process does not necessarily inform teachers about the collaborative product, (b) students' biases and/or reluctance in assessing peers within their group, and (c) there are no agreed practices or formulas to convert a group score or grade into individual scores or grades. However, since students seem to be able to assess peers validly in a formative context, formative group assessment combined with formative intra-group peer assessment might result in more construct valid measures compared to group assessment alone. Hence, this method for collaborative learning assessment might result in optimal feedback on students' learning process. With regard to summative purposes, the challenges with this method might cause group assessment combined with intra-group peer assessment to be invalid and lead to, for example, students not receiving their desired grade (Nordberg 2006) or a grade that does not represents their individual contribution to the group due to their dependence on their peers' domain-specific abilities.

Behavioural (mis)alignment of group assessment combined with intra-group peer assessment

Whether group assessment combined with intra-group peer assessment enhances the alignment of student behaviour largely depends on whether the peer assessment is valid, which may be more likely in formative than in summative contexts (Sridharan, Tai, and Boud 2019). When students know that their fellow group members will not 'reward' or 'punish' them for their contribution to the process and/or product, individual accountability is still lacking and student behaviour might be similar to that in the case of group assessment, and, for example, lead to free-riding behaviour. In addition, because group assessment typically concerns the collaborative product and intra-group peer assessment typically concerns the collaborative process, students might - as in the case of group assessment - adopt a performance-oriented approach rather than a learning-oriented approach (Pitt 2000). For example, a high-achiever might take over tasks of

low-achievers (Kagan 1995; Pitt 2000), and because both the high-achiever and the low-achiever benefit from this approach (from a performance-oriented perspective), students might decide to not validly conduct the intra-group peer assessment (Pitt 2000). Hence, when students do not validly assess each other because of (a) biases (e.g. because of friendship) or (b) a performance-oriented approach, group assessment combined with intra-group peer assessment might still lead to student behaviours that do not align with the objectives of collaborative learning (e.g. unequal task distribution).

When intra-group peer assessment is valid, which may be more likely in formative contexts (Sridharan, Tai, and Boud 2019), an ideal scenario in terms of aligned student behaviour might be as follows: positive interdependence is fostered by group assessment, whereas individual accountability is fostered by intra-group peer assessment. In other words, intra-group peer assessment might make students aware not only of what their peers should have done or failed to do but also what they themselves should have done or failed to do. For example, when students are unsatisfied with the behaviour or performance of specific group members, they can weigh that into their peer assessment. Likewise, the peer assessment by their fellow group members might prompt them to invest more effort and/or raise the quality of their contributions to the collaborative process and/or product. This heightened awareness of roles, responsibilities, and tasks might lead to student behaviour that is more aligned with the objectives of collaborative learning. As such, intra-group peer assessment might then help and stimulate the participation of students in groups and reduce free-riding behaviour (Johnston and Miles 2004; Shiu et al. 2012; Strijbos 2016; Sridharan, Muttakin, and Mihret 2018). Yet, even in summative contexts, intra-group peer assessment might be capable of fostering individual accountability. That is, when students do not know whether their peers will assess them validly, the 'threat' of being assessed 'accurately' by peers alone might lead to student behaviour better aligned with the objectives of collaborative learning; for example, more explicit communication and coordination of their own and their fellow group members' tasks and responsibilities.

In conclusion, group assessment combined with intra-group peer assessment might lead to student behaviour that is more aligned with the objectives of collaborative learning compared to group assessment and individual assessment. However, this is more likely in the case of a formative purpose because it might result in more valid intra-group peer assessments which in turn enhances students' sense of individual accountability. In the case of a summative purpose, in which students deliberately choose and/or might know that their peers will not validly assess their contribution to the collaborative process and/or product, students might still show misaligned behaviour such as free-riding and unequal task distribution.

Discussion

Collaborative learning is frequently employed in higher education (Jin 2012; Flores et al. 2015) and can be used for developing cognitive outcomes (e.g. knowledge), social outcomes (e.g. communication and collaboration skills), and motivational outcomes (e.g. attitudes) (Strijbos 2011). It has been shown that collaborative learning can lead to learning benefits compared to, for example, individual learning (Kyndt et al. 2013; Janssen 2014). However, since students differ in their domain-specific abilities (e.g. knowledge and/or skills) and the effort they invest into the collaboration, the assessment of collaborative learning is often considered as challenging by teachers (Lejk, Wyvill, and Farrow 1997; Ross, Rolheiser, and Hogaboam-Gray 1998; Ahern 2007; Gillies and Boyle 2010; Forslund Frykedal and Hammar Chiriatic 2011; Hammar Chiriatic and Granström 2012; de Hei et al. 2015; de Hei, Sjoer et al. 2016). Moreover, in comparison to other design components of collaborative learning such as 'group constellation' (de Hei, Strijbos et al. 2016), relatively little is known about how collaborative learning could be assessed (Strijbos 2011, 2016).

Table 1. Overview of the challenges with the assessment of collaborative learning.

	Construct validity and consequences		Behavioural (mis)alignment and consequences	
	Formative	Summative	Formative	Summative
Group assessment of collaborative learning	Given the focus of current curricula on measuring individual abilities, group assessment might be invalid.	Less able students might not pass a course while individually not meeting the requirements and standards. More able students might not pass a course or pass with a lower than the desired grade and/or a grade that does not represent their individual abilities.	Group assessment might elicit aligned and misaligned student behaviour.	
<i>Formative and summative</i> Neglects individual accountability Fosters positive interdependence	Feedback might not correspond to students' individual learning process.		Aligned: Group assessment fosters student's dependency on each other in order to succeed in the collaborative assignment, which can lead to, for example, students helping each other. Misaligned: Group assessment might lead to free-riding, social loafing, and the sucker effect. In addition, especially when the stakes are high, group assessment might result in groups adopting a performance-oriented approach rather than a learning-oriented approach.	
Individual assessment of collaborative learning	Given the focus of current curricula on measuring individual abilities, individual assessment is assumed to be more valid than group assessment although the assessment might still be affected by the collaborative setting.	Summative individual assessment of collaborative learning is, compared to group assessment, less likely to result in students passing a course who individually do not meet the required abilities; and less likely to more able students not passing a course or passing with a lower than the desired grade and/or a grade that does not represent their individual abilities.	Individual assessment of collaborative learning might elicit aligned and misaligned student behaviour.	
<i>Formative and summative</i> Fosters individual accountability Neglects positive interdependence	Feedback might better correspond to students' individual learning process compared to group assessment.		Aligned: Individual assessment of collaborative learning counteracts group assessments' problems with free-riding, social loafing, and the sucker-effect. Misaligned: Individual assessment of collaborative learning does not function as an incentive for students to engage in genuine collaboration. Even more than with group assessment, all it takes to succeed in a collaborative learning assignment is that students arrange superficial and high-level organisational matters, such as dividing subtasks, and decide who combines the separate pieces.	
Group assessment combined with intra-group peer assessment of collaborative learning	Given the focus of current curricula on measuring individual abilities, formative group assessment combined with intra-group peer assessment might increase the construct validity of group assessment.	Given the focus of current curricula on measuring individual abilities, summative group assessment combined with intra-group peer assessment might be invalid.	Group assessment and intra-group peer assessment of collaborative learning might elicit aligned student behaviour.	Group assessment and intra-group peer assessment of collaborative learning might elicit aligned and misaligned student behaviour.
<i>Formative</i> Fosters individual accountability Fosters positive				

(continued)



Table 1. Continued.

	Construct validity and consequences		Behavioural (mis)alignment and consequences	
	Formative	Summative	Formative	Summative
interdependence	Feedback might better correspond to students' individual learning process compared to group assessment.	Less able students passing a course while individually not meeting the requirements and standards. More able students not passing a course or passing with a lower than the desired grade and/or a grade that does not represent their individual abilities.	<u>Aligned:</u> Formative group assessment combined with intra-group peer assessment fosters student's dependency on each other in order to succeed in the collaborative assignment, which can lead to, for example, students helping each other. In addition, it might counteract group assessments' problems with free-riding, social loafing, and the sucker-effect.	<u>Aligned:</u> Summative group assessment combined with intra-group peer assessment fosters student's dependency on each other in order to succeed in the collaborative assignment, which can lead to, for example, students helping each other. <u>Misaligned:</u> Summative group assessment combined with intra-group peer assessment might lead to free-riding, social loafing, and the sucker effect. In addition, especially when the stakes are high (e.g., summative assessment), group assessment might result in groups adopting a performance-oriented approach rather than a learning-oriented approach.
<i>Summative</i> Neglects individual accountability Fosters positive interdependence				

To enhance teachers', researchers', and students' awareness for and need to develop what we refer to as 'collaborative learning assessment literacy', we discussed the two challenges of (a) ensuring construct validity and (b) mitigating misaligned student behaviour in relation to three frequently used and discussed methods for assessing collaborative learning - group assessment, individual assessment, and group assessment combined with intra-group peer assessment - with specific attention to the purpose of assessment (i.e. formative and summative). [Table 1](#) presents a (somewhat simplified) summary of the two challenges discussed in the preceding sections.

Several factors should be considered when interpreting these challenges. Firstly, they arise and exist because curricula and certification in higher education currently nearly exclusively stress the assessment of individual domain-specific abilities to decide whether students earn a degree (Boud, Cohen, and Sampson 1999; Nordberg 2006; Sharp 2006). When curricula and certification would allow and/or comprise different foci, the two challenges discussed in this paper could be different or ultimately even obsolete. As Payne et al. (2006, 446) pointed out:

Imagine for a moment if group work were the traditional and customary teaching style for students. Then, what would happen if we pulled the group out from under them and asked them to do an individual project? Many students would likely have the same aversion that they now have to group work. They'd have to learn new strategies of doing things, and this is something many of us resist.

Secondly, the two challenges we discussed should not be seen as fixed issues that occur in every group, in every class, and in every collaborative learning situation to an equal extent. Their occurrence can be very complex: this is determined by many factors such as actual student behaviour, student perceptions of the construct validity, whether students adopt a performance or learning oriented approach, and actual student differences in their domain-specific abilities. In addition, these factors also interact with each other. For example, the construct validity of assessment of collaborative learning influences the behaviour of students (e.g. invalid assessment might elicit a performance-oriented approach to collaborative learning), but how students behave also influences the construct validity of assessment of collaborative learning (e.g. a performance-oriented to collaborative learning approach might lead to an invalid assessment).

Thirdly, assessment is only one of the eight collaborative learning design components of collaborative learning (cf. the GLAID framework; de Hei, Strijbos et al. 2016). Hence, the design of assessment of collaborative learning and its alignment with the other collaborative learning design components also determines whether (a) the construct validity is compromised and (b) (mis)aligned student behaviour is likely to occur and to what degree. For example, when a course does not aim for the development of specific cognitive outcomes, but for preparing students for workplace situations in which a strong task specification within teams is required (e.g. surgical teams in which each individual in the operating room has a very specific task: surgeon, nurse, anaesthesiologist, etc.), individual assessment of collaborative learning might not lead to misaligned student behaviour since it aligns with the objectives of that course (relevant GLAID component: learning objectives and outcomes). Likewise, if the course aims for preparing students for intensive collaboration and integration of each individual's tasks, the potential impact of misaligned student behaviour should not be brushed away and justified as 'resembling possible unfair work practices' (Alden 2011), but instead be addressed through collaborative learning assessment and thereby ensure constructive alignment (Biggs 1996) and fair classroom assessment (Kagan 1995). In fact, a group assessment might be a construct valid assessment measure when all students in a group have similar levels of domain-specific abilities (relevant GLAID component: group constellation), work well together, and also explicitly document and acknowledge that they considered each other as equal contributors in the process and product throughout the entire collaboration. A case in point is Ohaja, Dunlea, and Muldoon (2013) who reported that student perceptions of a group assessment depended on the group dynamics within their group.

Implications for practice

Transition in collaborative learning activities and its assessment?

The first implication stems from the observation that challenges with the assessment of collaborative learning relate to, among other factors, the design of collaborative learning and the alignment of assessment with other collaborative learning design components. Hence, teachers should critically reflect on (a) their current assessment practices using the information in this paper and (b) the design of collaborative learning that they use in their course(s). Since there is no 'one-size-fits-all' design for collaborative learning, the GLAID framework of de Hei, Srijbos et al. (2016) could be a helpful tool to guide the design. The framework stimulates teachers to align, for example, their design choices with respect to the 'learning objectives and outcomes' with the 'assessment', 'structuring', and 'guidance' of collaborative learning within courses. It is not to say that all assessment challenges will be automatically solved by using this framework, but applying the framework will foster teachers' awareness of the role and implications of collaborative learning assessment and that it is not disconnected from, for example, the structuring of collaborative learning and guidance by the teacher.

Hence, more thoughtful design of collaborative learning and its assessment might lead to a change in the kind of collaborative learning activities that students participate in. For example, if the teachers' learning objective is to develop students' skills (i.e. communication and collaboration skills), there may be better methods than putting students together in a group and assuming that skills will develop without any guidance and/or formative and summative assessment of it (Felder and Brent 2001; Williams and Anderson 2008; Rebollar et al. 2010). As was pointed out by Johnson and Johnson (1999), a prerequisite for students learning together is that they are proficient in collaboration and communication; and if not, that they should be taught. To train students, lectures (e.g. for expanding knowledge) and interactive formats to practice skills (e.g. role-plays and simulations) are very suited.

For example, Rebollar et al. (2010) aimed to address students' shortcomings concerning collaboration in preparation for a collaborative learning assignment. To this end, students received ten theory lectures about collaboration and project management. Two seminars or workshops followed in which (a) instructors and former students shared experiences with collaborative learning, and (b) a psychologist lectured about group dynamics after which role-playing exercises were performed. Such interactive training activities (i.e. role-playing) allow for formative and summative individual assessment of collaboration knowledge and skills. This can be beneficial for the construct validity, while still adhering to the objective of developing collaboration skills. In addition, a training program as described by Rebollar et al. (2010) might better prepare students for collaborative learning outside of class (e.g. preparation for the future workplace). For example, it teaches students behaviour that teachers would like to see during a collaborative assignment, how to overcome collaboration struggles in the group, and how to assess fellow students' performances and ways to communicate the assessment to them (i.e. intra-group peer assessment). Not only might this directly lead to the development of collaboration skills and student behaviour that is more aligned with the objectives of collaborative learning, the improved collaboration and communication between students might also lead to higher construct validity of, for example, group assessment and group assessment combined with intra-group peer-assessment.

With respect to the learning objective of higher (cognitive) learning outcomes by fostering collaboration among students, it is questionable whether assessment is required to elicit these behaviours (Opdecam and Everaert 2018). Assessment of collaborative learning might even elicit student behaviour contradictory to these objectives (Kagan 1995; Pitt 2000; Hayek et al. 2015). On the one hand, it is generally known that summative assessment has a strong influence on the learning process and outcomes of students (Boud, Cohen, and Sampson 1999) and can be an important motivator for students (Macdonald 2003). On the other hand, formative assessment

might be more effective for fostering collaboration and consequently achieving higher (cognitive) learning outcomes. For example, one could imagine letting groups of students (a) prepare and present a presentation followed by comments from the teacher or fellow students (Kagan 1995) or (b) work together in tutorial groups while the teacher monitors the collaboration and provides feedback (Opdecam et al. 2014).

Hence, thoughtfully designing and aligning all collaborative learning design components might lead to collaborative learning settings that better adhere to the (intended) objectives and outcomes of collaborative learning. Moreover, enhanced collaborative learning design can also improve the construct validity of the collaborative learning assessment (e.g. by using formative or summative individual assessment and better training of students for collaboration out of class) and mitigate misaligned student behaviour (e.g. by training students in the desired collaborative behaviours and by teachers monitoring the collaboration in groups).

Students' transcripts

One could argue that the effect of receiving, for example, a group assessment as a result of a collaborative learning assignment has a negligible impact on students' transcripts, because individual assessment of individual assignments (e.g. written tests, term papers) typically forms the larger part of students' overall grades for any given course. However, even if we were to assume that, for example, a group grade is a small factor in the overall course assessment and/or an educational programme, it can still have significant consequences for students (Kagan 1995), such as (not) passing a course or (not) graduating with distinction. In addition, as collaborative learning is increasingly used in higher education (Jin 2012; Flores et al. 2015) it is reasonable to assume that the share of collaborative learning assessment in educational programmes increases as well.

Therefore, the second implication entails the open question whether it is necessary to represent and differentiate in students' transcripts between grades that they achieved individually or collaboratively. That is, although most contemporary curricula comprise many courses that include collaborative learning to some degree, the outcome of students' contribution to collaborative learning is typically not specified separately as a partial grade that appears on the transcript. Instead, the grade that appears on the transcript is either only the result of a group assessment or the group assessment outcome is mixed with an individual grade on a separate assignment or test that is disconnected from the group assignment. Hence, it is currently unclear which grades were truly achieved individually or collaboratively. Although separating individual and group grades will not necessarily be a solution to the challenges discussed in this paper (e.g. students might still adopt a performance-oriented approach or free-ride), grades as a result of collaborative learning will at least not be misinterpreted as a reflection of individual domain-specific abilities which might increase the construct validity of, for example, a group assessment. In addition, it will provide insight in what a student can do individually and in collaboration with others which provides, for example, admission committees and employers more detailed information about the domain-specific abilities of future students or employees.

Implications for research

In our view, future research on the assessment of collaborative learning in higher education should focus on two key themes. Firstly, more knowledge about the methods available for the assessment of collaborative learning should be gained. Currently, there is no large-scale overview and analysis of (a) (validated) methods and instruments that teachers in higher education can use for assessing collaborative learning and (b) teachers' actual rationale for and practices of the assessment of collaborative learning in higher education. The latter is especially relevant for the

discussion on how much impact grades as a result of collaborative learning currently have or should have on students' transcripts. For example, future research could analyse whether and to what extent collaborative learning occurs in educational programmes and the degree to which its assessment has an impact on students' course grades and transcript of records.

Secondly, more empirical research should be done to better understand the relation between several choices in the collaborative learning design components (e.g. homogeneity/heterogeneity in student groups, training in collaborative learning skills, monitoring and feedback by teachers) and the construct validity and (mis)aligned student behaviour of different assessment methods of collaborative learning (e.g. group assessment, individual assessment, and group assessment combined with peer assessment; as well as formative or summative purpose). For example, what is the construct validity of group scores or grades in student groups that are (more) homogeneous in individual domain-specific abilities (e.g. knowledge and skills) compared to (more) heterogeneous student groups? What are the consequences of (mis)aligned student behaviour (e.g. free-riding behaviour and adopting a performance-oriented approach) when teachers monitor student groups compared to when no monitoring is applied? By empirically studying what effects collaborative learning design choices have on both construct validity and (mis)aligned student behaviour, more specific guidelines can be formulated to develop 'collaborative learning assessment literacy' and support teachers in designing collaborative learning activities and their assessment.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Hajo Meijer received a BA and MA in psychology at the University of Groningen. He is PhD student in the Department of Educational Sciences; University of Groningen in the Netherlands. His PhD research focuses on the assessment of collaborative learning in higher education.

Rink Hoekstra is an assistant professor at the Department of Educational Sciences; University of Groningen in the Netherlands. His research mainly focuses on how researchers understand and use statistical outcomes, and how the scientific context influences why they use it in a certain way. He has published regularly about use and understanding of significance testing and its alternatives.

Jasperina Brouwer is assistant professor at the Department of Educational Sciences; University of Groningen in the Netherlands. She investigates the impact of social networks and social capital on academic and professional development within the educational context and at the workplace.

Jan-Willem Strijbos is a full professor of learning and instruction in the Department of Educational Sciences; University of Groningen in the Netherlands. His research focuses on the design, implementation, and effectiveness of interactive learning practices, for example, collaborative learning, peer assessment and feedback, learning communities—in physical and virtual settings.

ORCID

Hajo Meijer  <http://orcid.org/0000-0002-8563-5333>

Rink Hoekstra  <http://orcid.org/0000-0002-1588-7527>

Jasperina Brouwer  <http://orcid.org/0000-0002-7332-9320>

Jan-Willem Strijbos  <http://orcid.org/0000-0003-4180-688X>

References

Ahern, A. 2007. "What Are the Perceptions of Lecturers towards Using Cooperative Learning in Civil Engineering?" *European Journal of Engineering Education* 32 (5): 517–526. doi:10.1080/03043790701433152.

- Alden, J. 2011. "Assessment of Individual Student Performance in Online Team Projects." *Journal of Asynchronous Learning Networks* 15 (3): 5–20. doi:10.24059/olj.v15i3.193.
- Almond, R. J. 2009. "Group Assessment: Comparing Group and Individual Undergraduate Module Marks." *Assessment & Evaluation in Higher Education* 34 (2): 141–148. doi:10.1080/02602930801956083.
- Augar, N., C. J. Woodley, D. Whitefield, and M. Winchester. 2016. "Exploring Academics' Approaches to Managing Team Assessment." *International Journal of Educational Management* 30 (6): 1150–1162. doi:10.1108/IJEM-06-2015-0087.
- Bacon, D. R. 2005. "The Effect of Group Projects on Content-Related Learning." *Journal of Management Education* 29 (2): 248–267. doi:10.1177/1052562904263729.
- Bacon, D. R., K. A. Stewart, and W. S. Silver. 1999. "Lessons from the Best and Worst Student Team Experiences: How a Teacher Can Make the Difference." *Journal of Management Education* 23 (5): 467–488. doi:10.1177/105256299902300503.
- Bacon, D. R., K. A. Stewart, and S. Stewart-Belle. 1998. "Exploring Predictors of Student Team Project Performance." *Journal of Marketing Education* 20 (1): 63–71. doi:10.1177/027347539802000108.
- Barfield, R. L. 2003. "Students' Perceptions of and Satisfaction with Group Grades and the Group Experience in the College Classroom." *Assessment & Evaluation in Higher Education* 28 (4): 355–370. doi:10.1080/0260293032000066191.
- Biggs, J. 1996. "Enhancing Teaching through Constructive Alignment." *Higher Education* 32 (3): 347–364. doi:10.1007/BF00138871.
- Bolton, M. K. 1999. "The Role of Coaching in Student Teams: A "Just-in-Time" Approach to Learning." *Journal of Management Education* 23 (3): 233–250. doi:10.1177/105256299902300302.
- Boud, D., R. Cohen, and J. Sampson. 1999. "Peer Learning and Assessment." *Assessment & Evaluation in Higher Education* 24 (4): 413–426. doi:10.1080/0260293990240405.
- Brookhart, S. M. 2013. *Grading and Group Work: How Do I Assess Individual Learning When Students Work Together?* Alexandria, VA: ASCD.
- Bushell, G. 2006. "Moderation of Peer Assessment in Group Projects." *Assessment & Evaluation in Higher Education* 31 (1): 91–108. doi:10.1080/02602930500262395.
- Carson, K. M., and R. E. Glaser. 2010. "Chemistry is in the News: Assessing Intra-Group Peer Review." *Assessment & Evaluation in Higher Education* 35 (4): 381–402. doi:10.1080/02602930902862826.
- Carver, T. L., and A. Stickley. 2012. "Teamwork in First Year Law Units: Can It Work?" *Journal of University Teaching & Learning Practice* 9 (2): 2–32. <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1265&context=jutlp>.
- Cheng, W., and M. Warren. 2000. "Making a Difference: Using Peers to Assess Individual Students' Contributions to a Group Project." *Teaching in Higher Education* 5 (2): 243–255. doi:10.1080/135625100114885.
- Conway, R., D.Kember, A. Sivan, and M. Wu. 1993. "Peer Assessment of an Individual's Contribution to a Group Project." *Assessment & Evaluation in Higher Education* 18 (1): 45–56. doi:10.1080/0260293930180104.
- De Hei, M., J. W. Strijbos, E. Sjoer, and W. Admiraal. 2016. "Thematic Review of Approaches to Design Group Learning Activities in Higher Education: The Development of a Comprehensive Framework." *Educational Research Review* 18: 33–45. doi:10.1016/j.edurev.2016.01.001.
- De Hei, M. S. A., E. Sjoer, W. Admiraal, and J. W. Strijbos. 2016. "Teacher Educators' Design and Implementation of Group Learning Activities." *Educational Studies* 42 (4): 394–409. doi:10.1080/03055698.2016.1206461.
- De Hei, M. S. A., J. W. Strijbos, E. Sjoer, and W. Admiraal. 2015. "Collaborative Learning in Higher Education: Lecturers' Practices and Beliefs." *Research Papers in Education* 30 (2): 232–247. doi:10.1080/02671522.2014.908407.
- Dijkstra, J., M. Latijnhouwers, A. Norbart, and R. A. Tio. 2016. "Assessing the "I" in Group Work Assessment: State of the Art and Recommendations for Practice." *Medical Teacher* 38 (7): 675–682. doi:10.3109/0142159X.2016.1170796.
- Diprose, M. F., P. L. Judd, and A. S. Morris. 1997. "The Sheffield Industrial Project Scheme: Team Teaching for Engineering Students." *Teaching in Higher Education* 2 (1): 59–68. doi:10.1080/1356251970020105.
- Felder, R. M., and R. Brent. 2001. "Effective Strategies for Cooperative Learning." *Journal of Cooperation & Collaboration in College Teaching* 10 (2): 69–75. <http://www.academia.edu/download/39262475/CASTLLawrieDocs.pdf>.
- Flores, M. A., A. M. Veiga Simão, A. Barros, and D. Pereira. 2015. "Perceptions of Effectiveness, Fairness and Feedback of Assessment Methods: A Study in Higher Education." *Studies in Higher Education* 40 (9): 1523–1534. doi:10.1080/03075079.2014.881348.
- Forsell, J. K., E. Forslund Frykedal, and E. Hammar Chiriac. 2020. "Group Work Assessment: Assessing Social Skills at Group Level." *Small Group Research* 51 (1): 87–124. doi:10.1177/1046496419878269.
- Forslund Frykedal, K., and E. Hammar Chiriac. 2011. "Assessment of Students' Learning When Working in Groups." *Educational Research* 53 (3): 331–345. doi:10.1080/00131881.2011.598661.
- Gibbs, G. 1995. *Learning in Teams: A Tutor Guide*. 1st ed. Oxford: Oxford Centre for Staff Development.
- Gillies, R. M. 2007. *Cooperative Learning: Integrating Theory and Practice*. Thousand Oaks, CA: Sage.

- Gillies, R. M., and M. Boyle. 2010. "Teachers' Reflections on Cooperative Learning: Issues of Implementation." *Teaching and Teacher Education* 26 (4): 933–940. doi:10.1016/j.tate.2009.10.034.
- Goldfinch, J. 1994. "Further Developments in Peer Assessment of Group Projects." *Assessment & Evaluation in Higher Education* 19 (1): 29–35. doi:10.1080/0260293940190103.
- Hammar Chiriach, E., and K. Granström. 2012. "Teachers' Leadership and Students' Experience of Group Work." *Teachers and Teaching* 18 (3): 345–363. doi:10.1080/13540602.2012.629842.
- Harvey, H., J. Keen, C. Robinson, J. Roff, and T. Gross. 2019. "Quantitative Analysis of Approaches to Group Marking." *Assessment & Evaluation in Higher Education* 44 (8): 133–1147. doi:10.1080/02602938.2019.1572069.
- Hastie, C. R. 2018. "TeamUP': An Approach to Developing Teamwork Skills in Undergraduate Midwifery Students." *Midwifery* 58: 93–95. doi:10.1016/j.midw.2017.12.026.
- Hayek, A.-S., C. Toma, D. Oberlé, and Fabrizio Butera. 2015. "Grading Hampers Cooperative Information Sharing in Group Problem Solving." *Social Psychology* 46 (3): 121–131. doi:10.1027/1864-9335/a000232.
- Janssen, J. 2014. "Opening the Black Box of Collaborative Learning: A Meta-Analysis Investigating the Antecedents and Consequences of Collaborative Interaction." <http://www.nwo.nl/actueel/nieuws/2014/magw/wat-samenwerkend-leren-effectief-maakt.htm>.
- Jin, X. H. 2012. "A Comparative Study of Effectiveness of Peer Assessment of Individuals' Contributions to Group Projects in Undergraduate Construction Management Core Units." *Assessment & Evaluation in Higher Education* 37 (5): 577–589. doi:10.1080/02602938.2011.557147.
- Johnson, D. W. 1981. "Student-Student Interaction: The Neglected Variable in Education." *Educational Researcher* 10 (1): 5–10. doi:10.3102/0013189X010001005.
- Johnson, D. W., and R. T. Johnson. 1999. "Making Cooperative Learning Work." *Theory into Practice* 38 (2): 67–73. doi:10.1080/00405849909543834.
- Johnston, L., and L. Miles. 2004. "Assessing Contributions to Group Assignments." *Assessment & Evaluation in Higher Education* 29 (6): 751–768. doi:10.1080/0260293042000227272.
- Kagan, S. 1995. "Group Grades Miss the Mark." *Educational Leadership* 52 (8): 68–71. https://www.researchgate.net/profile/Spencer_Kagan/publication/234597096_Group_Grades_Miss_the_Mark/links/58cc7f834585157b6dac119d/Group-Grades-Miss-the-Mark.pdf.
- Kaufman, J. H., and C. D. Schunn. 2011. "Students' Perceptions about Peer Assessment for Writing: Their Origin and Impact on Revision Work." *Instructional Science* 39 (3): 387–406. doi:10.1007/s11251-010-9133-6.
- Knight, J. 2004. "Comparison of Student Perception and Performance in Individual and Group Assessments in Practical Classes." *Journal of Geography in Higher Education* 28 (1): 63–81. doi:10.1080/0309826042000198648.
- Kyndt, E., E. Raes, B. Lismont, F. Timmers, E. Cascallar, and F. Dochy. 2013. "A Meta-Analysis of the Effects of Face-to-Face Cooperative Learning: Do Recent Studies Falsify or Verify Earlier Findings?" *Educational Research Review* 10: 133–149. doi:10.1016/j.edurev.2013.02.002.
- Lejk, M., and M. Wyvill. 2001. "Peer Assessment of Contributions to a Group Project: A Comparison of Holistic and Category-Based Approaches." *Assessment & Evaluation in Higher Education* 26 (1): 61–72. doi:10.1080/02602930020022291.
- Lejk, M., M. Wyvill, and S. Farrow. 1999. "Group Assessment in Systems Analysis and Design: A Comparison of the Performance of Streamed and Mixed-Ability Groups." *Assessment & Evaluation in Higher Education* 24 (1): 5–14. doi:10.1080/0260293990240101.
- Lejk, M., M. Wyvill, and S. Farrow. 1997. "Group Learning and Group Assessment on Undergraduate Computing Courses in Higher Education in the UK: Results of a Survey." *Assessment & Evaluation in Higher Education* 22 (1): 81–91. doi:10.1080/0260293970220106.
- Li, L. 2001. "Some Refinements on Peer Assessment of Group Projects." *Assessment & Evaluation in Higher Education* 26 (1): 5–18. doi:10.1080/0260293002002255.
- Macdonald, J. 2003. "Assessing Online Collaborative Learning: Process and Product." *Computers & Education* 40 (4): 377–391. doi:10.1016/S0360-1315(02)00168-9.
- Magin, D. 2001. "Reciprocity as a Source of Bias in Multiple Peer Assessment of Group Work." *Studies in Higher Education* 26 (1): 53–63. doi:10.1080/03075070123503.
- Messick, S. 1995. "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning." *American Psychologist* 50 (9): 741–749. doi:10.1037/0003-066X.50.9.741.
- Moore, P., G. Hampton. 2015. "It's a Bit of a Generalisation, but...': Participant Perspectives on Intercultural Group Assessment in Higher Education." *Assessment & Evaluation in Higher Education* 40 (3): 390–406. doi:10.1080/02602938.2014.919437.
- Nepal, K. P. 2012. "An Approach to Assign Individual Marks from a Team Mark: The Case of Australian Grading System at Universities." *Assessment & Evaluation in Higher Education* 37 (5): 555–562. doi:10.1080/02602938.2011.555815.
- Neus, J. N. 2011. "Peer Assessment Accounting for Student Agreement." *Assessment & Evaluation in Higher Education* 36 (3): 301–314. doi:10.1080/02602930903342315.

- Nordberg, D. 2008. "Group Projects: More Learning? Less Fair? a Conundrum in Assessing Postgraduate Business Education." *Assessment & Evaluation in Higher Education* 33 (5): 481–492. doi:10.2139/ssrn.887145.
- Nordberg, D. 2006. Fairness in Assessing Group Projects: A Conceptual Framework for Higher Education. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=873605. doi:10.2139/ssrn.873605.
- Ohaja, M., M. Dunlea, and K. Muldoon. 2013. "Group Marking and Peer Assessment during a Group Poster Presentation: The Experiences and Views of Midwifery Students." *Nurse Education in Practice* 13 (5): 466–470. doi:10.1016/j.nepr.2012.11.005.
- Onyia, O. P. 2014. "Integrating Teacher-and Peer-Assessments of Group Coursework Assignments in Business Education: Some Innovative Methods." *Research in Higher Education Journal* 22: 1–10.
- Opdecam, E., and P. Everaert. 2018. "Seven Disagreements about Cooperative Learning." *Accounting Education* 27 (3): 223–233. doi:10.1080/09639284.2018.1477056.
- Opdecam, E., P. Everaert, H. Van Keer, and F. Buyschaert. 2014. "Preferences for Team Learning and Lecture-Based Learning among First-Year Undergraduate Accounting Students." *Research in Higher Education* 55 (4): 400–432. doi:10.1007/s11162-013-9315-6.
- Payne, B. K., E. Monk-Turner, D. Smith, and M. Sumter. 2006. "Improving Group Work: Voices of Students." *Education* 126 (3): 441–448. https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1023&context=sociology_criminal-justice_fac_pubs.
- Pitt, M. J. 2000. "The Application of Games Theory to Group Project Assessment." *Teaching in Higher Education* 5 (2): 233–241. doi:10.1080/135625100114876.
- Plastow, N., G. Spiliotopoulou, and S. Prior. 2010. "Group Assessment at First Year and Final Degree Level: A Comparative Evaluation." *Innovations in Education and Teaching International* 47 (4): 393–403. doi:10.1080/14703297.2010.518427.
- Quilter, S. M., and J. K. Gallini. 2000. "Teachers' Assessment Literacy and Attitudes." *The Teacher Educator* 36 (2): 115–131. doi:10.1080/0887873000955257.
- Rebollar, R., I. Lidón, J. L. Cano, F. Gimeno, and P. Qvist. 2010. "A Tool for Preventing Teamwork Failure: The TFP Questionnaire." *International Journal of Engineering Education* 26 (4): 784–799.
- Roberts, T. S., and J. M. McInerney. 2007. "Seven Problems of Online Group Learning (and Their Solutions)." *Journal of Educational Technology & Society* 10 (4): 257–268. https://www.researchgate.net/profile/Tim_Roberts2/publication/220374945_Seven_Problems_of_Online_Group_Learning_and_Their_Solutions/links/0912f51019f3bda785000000.pdf.
- Roskams, T. 1999. "Chinese EFL Students' Attitudes to Peer Feedback and Peer Assessment in an Extended Pairwork Setting." *RELC Journal* 30 (1): 79–123. doi:10.1177/00368829903000105.
- Ross, J. A., C. Rolheiser, and A. Hogaboam-Gray. 1998. "Student Evaluation in Co-Operative Learning: Teacher Cognitions." *Teachers and Teaching* 4 (2): 299–316. doi:10.1080/1354060980040207.
- Salomon, G., and T. Globerson. 1989. "When Teams Do Not Function the Way They Ought to." *International Journal of Educational Research* 13 (1): 89–99. doi:10.1016/0883-0355(89)90018-9.
- Scriven, M. 1967. "The Methodology of Evaluation." In *Perspectives of Curriculum Evaluation*, edited by R. E. Stake, 39–55. Chicago: Rand McNally.
- Sharp, S. 2006. "Deriving Individual Student Marks from a Tutor's Assessment of Group Work." *Assessment & Evaluation in Higher Education* 31 (3): 329–343. doi:10.1080/02602930500352956.
- Shiu, A. T. Y., C. W. H. Chan, P. Lam, J. Lee, and A. N. L. Kwong. 2012. "Baccalaureate Nursing Students' Perceptions of Peer Assessment of Individual Contributions to a Group Project: A Case Study." *Nurse Education Today* 32 (3): 214–218. doi:10.1016/j.nedt.2011.03.008.
- Sivan, A. 2000. "The Implementation of Peer Assessment: An Action Research Approach." *Assessment in Education: Principles, Policy & Practice* 7 (2): 193–213. doi:10.1080/713613328.
- Slavin, R. E. 1980. "Cooperative Learning in Teams: State of the Art." *Educational Psychologist* 15 (2): 93–111. doi:10.1080/00461528009529219.
- Spatar, C., H. Penna, H. Mills, V. Kutija, and M. Cook. 2015. "A Robust Approach for Mapping Group Marks to Individual Marks Using Peer Assessment." *Assessment & Evaluation in Higher Education* 40 (3): 371–389. doi:10.1080/02602938.2014.917270.
- Sridharan, B., M. B. Muttakin, and D. G. Mihret. 2018. "Students' Perceptions of Peer Assessment Effectiveness: An Explorative Study." *Accounting Education* 27 (3): 259–285. doi:10.1080/09639284.2018.1476894.
- Sridharan, B., J. Tai, and D. Boud. 2019. "Does the Use of Summative Peer Assessment in Collaborative Group Work Inhibit Good Judgement?" *Higher Education* 77 (5): 853–870. doi:10.1007/s10734-018-0305-7.
- Srijbos, J. W. 2011. "Assessment of (Computer-Supported) Collaborative Learning." *IEEE Transactions on Learning Technologies* 4 (1): 59–73. doi:10.1109/TLT.2010.37.
- Srijbos, J. W. 2016. "Assessment of Collaborative Learning." In *Handbook of Social and Human Conditions in Assessment*, edited by G. T. L. Brown & L. Harris, 302–318. New York: Routledge.
- Srijbos, J. W., D. Sluijsmans, and K. Stegmann. 2017. "Inferring Individual Scores from Group Scores via Peer Assessment: Part 1 – Methodological Review." Paper Presented at the "Assessment and Evaluation SIG." Invited Symposium Conducted at the 17th Biennial EARLI Conference, Tampere, August-September, 29–2.

- Van Aalst, J. 2013. "Assessment in Collaborative Learning." In *The International Handbook of Collaborative Learning*, edited by C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, and A. M. O'Donnell, 280–293. New York, NY: Routledge
- Van den Bergh, V., D. Mortelmans, P. Spooren, P. Van Petegem, D. Gijbels, and G. Vanthournout. 2006. "New Assessment Modes within Project-Based Education-the Stakeholders." *Studies in Educational Evaluation* 32 (4): 345–368. doi:[10.1016/j.stueduc.2006.10.005](https://doi.org/10.1016/j.stueduc.2006.10.005).
- Webb, N. M. 1993. "Collaborative Group versus Individual Assessment in Mathematics: Processes and Outcomes." *Educational Assessment* 1 (2): 131–152. doi:[10.1037/e670282011-001](https://doi.org/10.1037/e670282011-001).
- Willcoxson, L. E. 2006. "It's Not Fair!": Assessing the Dynamics and Resourcing of Teamwork." *Journal of Management Education* 30 (6): 798–808. doi:[10.1177/1052562906287964](https://doi.org/10.1177/1052562906287964).
- Williams, S. K., and J. S. Anderson. 2008. "Teams Lab: Promoting Effective Teamwork in Operations Management Classes." *Decision Sciences Journal of Innovative Education* 6 (1): 159–166. doi:[10.1111/j.1540-4609.2007.00165.x](https://doi.org/10.1111/j.1540-4609.2007.00165.x).
- Wu, C., E. Chanda, and J. Willison. 2014. "Implementation and Outcomes of Online Self and Peer Assessment on Group Based Honours Research Projects." *Assessment & Evaluation of Higher Education* 39 (1): 21–37. doi:[10.1080/02602938.2013.779634](https://doi.org/10.1080/02602938.2013.779634).