Group Work Assessment: Assessing Social Skills at Group Level

Johan Forsell, Karin Forslund Frykedal and Eva Hammar Chiriac

Tweet

LIU LINKÖPING UNIVERSITY

Group Work Assessment: Assessing Social Skills at Group Level

Johan Forsell, Karin Forslund Frykedal and Eva Hammar Chiriac

Abstract

Group work assessment is often described by teachers as complex and challenging, with individual assessment and fair assessment emerging as dilemmas. The aim of this literature review is to explore and systematize research about group work assessment in educational settings. This is an integrated research area consisting of research combining group work and classroom assessment. A database search was conducted, inspired by the guidelines of the PRISMA. The analysis and categorization evolved into a typology consisting of five themes: (a) *purpose of group work assessment*, (b) *what is assessed in group work*, (c) *methods for group work assessment*, (d) *effects and consequences of group work assessment*, and (e) *quality in group work assessment*. The findings reveal that research in the field of group work assessment notably focuses on social skills and group processes. Peer assessment plays a prominent role and teachers as assessors are surprising absences in the reviewed research.

Keywords

Group work, assessment, systematic review, group work assessment, education

Group Work Assessment: Assessing Social Skills at Group Level

The goal of this review is to gain a holistic, systematic overview of the research concerning group work assessment. The review covers research from psychology and educational science focusing on the process of assessing group work in an educational setting. Group work assessment is seen in this review as a holistic concept covering a broad scope, including both formative and summative assessment, evaluation, grading, and a variety of methods.

Previous research on group work in educational settings has generated knowledge in several areas, including: (a) learning outcome (i.e., group work promotes both academic achievement and social skills/collaborative abilities, Baines, Blatchford, & Chowne, 2007; Gillies & Boyle, 2010, 2011; Hammar Chiriac, 2014); (b) pedagogical approaches (i.e., cooperative learning, collaborative learning, problem-based learning, Gillies, 2007; Davidson & Howell Major, 2014); (c) communication (i.e., the importance of language skills, Forslund Frykedal & Hammar Chiriac, 2011; Gillies, 2017; Gillies & Kahn, 2008; Webb, Franke, Ing, Turrou, & Johnson, 2015); and (d) other aspects influencing work and processes in groups (i.e., task, roles, ambition, trust, support, and context, Hammar Chiriac, 2003, 2008; Hammar Chriac & Granström, 2012, Forslund Frykedal, 2008; Forslund Frykedal & Hammar Chiriac, 2018). In sum, research has provided us with scientific knowledge about group work, primarily from the perspectives of group psychology and education. However, we still lack answers as to why some instances of group work turn out favourably while others do not. One aspect that might add some information could be classroom assessment.

In classroom assessment, knowledge has also been generated in several areas. McMillan (2012) divides research on classroom assessment into the following areas: (a) the

technical quality of classroom assessment, or its validity, reliability, and fairness (e.g., Alm & Colnerud, 2015; Black, Harrison, Hodgen, Marshall, & Serret, 2010; Harlen, 2005); (b) formative assessment, or gathering evidence of student understanding, feedback, and instructional correctives (e.g., Black & Wiliam, 1998; Harlen & James, 1997); (c) summative assessment, or the impact of summative assessment and tests, motivation for learning, and grading (e.g., Black et al., 2010; Harlen & James, 1997); (d) methods of classroom assessment, or, testing formats, such as constructed response, selected response, performance assessment, portfolios, student self-assessment, and peer assessment (e.g., Andrade, Du, & Wang, 2008; Haladyna, Downing, & Rodriguez, 2002; Rodriguez, 2005); and (e) differentiated classroom assessment, or demographics and differentiated classroom assessment, classroom assessment in special education, classroom assessment in different subjects, such as mathematics, social studies, science, and literacy (e.g., Hodgen & Marshall, 2005; Kamps, Wendland, Culpepper, Lamb, & Charter, 2006).

Combining and integrating group work and classroom assessment opens up the field of group work assessment. Previous research has shown scientific support for the interdependence between learning (e.g., appropriated in group work) and assessment (Black et al., 2003; Johnson & Johnson, 2004; van Aalst, 2013; Wiliam, 2009, 2011; Wiliam & Thompson, 2007). Hence, group work assessment is highly relevant when organizing group work in educational settings. Often group work assessment is described by teachers as a complex and challenging issue (Forslund Frykedal & Hammar Chiriac, 2011, 2016; Murray & Boyd, 2015). For instance, there are some challenges in structuring fair assessments (Gammie & Matson, 2007; Jin 2012; Onyia & Allen, 2012) or in gleaning information about who did what (Ko, 2013; Murray & Boyd, 2015). In our experience, group work assessment

is a rather neglected research area. Thus, there is little theoretical knowledge and few useful tools to assist teachers in resolving these conflicting demands.

The aim of this review is to explore, systematize, and compile research about group work assessment in educational settings, an integrated research area consisting of research combining group work and classroom assessment. From the collected articles, a typology can be created which identifies the range of this combined area and how its two parts interrelate. Furthermore, by identifying knowledge gaps, we also hope to suggest directions for profitable future research.

Method

Search Procedure

Group work assessment can be described as a bi-disciplinary and integrated research area, in which psychology and educational science are combined. It was necessary to use key words and concepts from both disciplines as search terms. To find relevant search terms, the thesaurus of the ERIC database was scanned. From this scan, two lists of search terms were assembled. For psychology, the search terms were *group work*, *cooperative learning*, *collaborative learning*, *cooperative education*, *peer teaching*, *group project*, and *group activities*. For educational sciences, the search terms were *assess**, *evaluation*, *measured*, *grad**, *credential*, *feedback*, and *rubric**. Since there are many different search terms for both group work and classroom assessment, a broad approach was taken regarding which terms to include. In some search terms, the stem of a word was used in order to cover variations.

Due to our cross-disciplinary approach, databases in both psychology and educational science were used. Accordingly, the search process continued in the two databases PsychINFO and ERIC. To ensure that Swedish research was included, the database Libris was also included. The database searches were conducted in April 2017. The following

limitations were imposed to exclude irrelevant research. The articles included in the review should be: (a) peer reviewed, (b) written in English or Swedish, and either (c) published in scientific journals, or (d) published as a dissertation. Further limitations were: (e) the articles or dissertations should focus on an educational setting, or (f) they should be about classroom assessment in a group context. With these limitations, the search was conducted by combining all the words of the search terms from both fields in all possible variations. No time interval was set for the search since we assumed that the amount of relevant research would be limited. Neither were any particular grades used, the search covered everything from primary school to college and university.

The search process was inspired by the guidelines and principles of the PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009) for conducting a systematic review. The first phase of the PRISMA statement, to identify records through database searching, was implemented in the PsychINFO, ERIC, and Libris databases; these searches returned 2600, 2375, and 135 search results, respectively; and all of these were screened. Initially, the screening procedure was pursued by scanning titles and keywords. If there was any ambiguity about a text's relevance, the abstracts were read. The second phase of the PRISMA statement, screening the results found in the search, also entails the exclusion of research that is not considered relevant. After this process, 4881 search results were excluded due to being considered irrelevant to the aims of this review; 229 publications were considered relevant and selected. Consequently. The third phase of the PRISMA statement involves deciding which research is eligible for the systematic review, and giving reasons for the exclusion of any non-eligible texts. This process led to the exclusion of 107 publications because they did not focus on an educational setting. After this process, there were 122 publications remaining that we considered relevant for the fourth phase of the PRISMA criteria. These articles were

selected and included in the analytical portion of the systematic review. During this phase, a thematic analysis of abstracts from the selected literature was implemented (see below), followed by an additional search to identify relevant key references found in these publications. Through this process, 21 supplementary articles were identified and added to the list of relevant articles for the review, which led to a total of 143 articles (Figure 1).

Process of Analysis

In order to organize the data, and to find patterns and themes, a thematic analysis was carried out. The purpose of thematic analysis is to arrange the dataset into themes followed by underlying subthemes by means of six phases (Braun & Clarke, 2006). A vast amount of literature was found during the search process, which consequently gave a plentiful volume of data to handle. To be able to organize this data, the idea was to build an initial structure by locating and thematising the main content of the abstracts.

The next step in a thematic analysis (Braun & Clarke, 2006) is to become familiar with the data. This familiarization was carried out by reading all 143 selected abstracts multiple times while noting initial ideas or patterns. Phase two consists of generating initial codes. The coding procedure was accomplished by inductively outlining central phrases, meanings of varying length, or words in the abstracts. Each outlined phrase, meaning, or word was labelled with a code that summarized or caught the essence of the outlining. Between two and 12 codes were generated for each abstract. Finally, all the data segments with corresponding codes were systematically copied into a single document. This last procedure was important because it kept the link between the code and the meaning unit intact for the next phase. The third phase entails a process of sorting codes based on differences and similarities in order to consider how a combination of similar codes can form a theme. This was achieved using a comparative analysis. During this phase, thematic maps,

including different codes, constructions of subthemes, and the writing process, were important tools. Five tentative themes emerged from this process, each with several subthemes. The codes were not seen as exclusive to only one theme; thus, some codes were included in more than one of the five themes. For example, peer assessment reoccurs in several themes but is enlightened through different angles. The comparative analysis continued with the fourth phase (Braun & Clark, 2006), which is described as reviewing the themes, with the objective that data within a theme should cohere together meaningfully. This is pursued by comparing codes within and between themes. During this process, some codes were moved, and others were deleted because they had no relevance to group work assessment in educational settings. What were initially tentative suggestions slowly emerged as more defined themes; thus, subthemes emerged more clearly through this process. In exceptional cases, where information in the abstracts was not sufficient to reach an understanding of the research content, relevant parts of the publications were read. Finally, during this fourth phase, a thematic map (including maps of the five themes and their subthemes) was constructed. This process also led to the exclusion of a further 53 articles and dissertations since they lacked relevance and were not about group work assessment. Consequently, 90 articles and dissertations remained in the dataset.

Braun and Clarke (2006) stipulate that researchers should consider the validity of their themes in relation to the data set. In our analysis, this consideration was part of the writing process, in which we continuously referred to the codes and their related meaning units when writing about the themes. The fifth phase, defining and naming the themes, was an ongoing process throughout the analysis, and the names of the themes were modified back and forth until they received a definitive name. The sixth phase is to produce the report. Here, the process continued with a systematic review of all the selected literature in full text. This

was a deductive process during which relevant information from the literature was placed into the structure based on the thematic map of themes and subthemes. During this process of writing, no main themes were added or deleted; however, some subthemes were slightly rearranged when this became necessary. Also, in some subthemes new aspects were added. The most extensive changes were made to the themes *what is assessed in group work?* and *methods for group work assessment*, which were rearranged because the number of details added made it necessary to reconstruct the themes.

Our experience was that the process of writing started early in the analysis and can be described as running in parallel with the process of building the thematic map. Henceforth in this review, the thematic map is labelled as a typology. Braun and Clarke (2006) stated that the analysis is not linear, which is consistent with our experience in that the process moved back and forth throughout the analysis alongside the writing of the themes. The themes were revised, reconstructed, and refined throughout the writing process. Furthermore, during the process of reading the literature in full text, the exclusion of publications continued as it became clear that some which had been previously identified lacked relevance for the review since they were not situated in the context of group work assessment. Across these steps, 90 full-text scholarly articles and dissertations were included. During the process of further reading, an additional seven publications were excluded; five of these were not situated in the context of group work, and two of the dissertations could not be found in full text. At the conclusion of the analysis process, 83 articles and dissertations were included in the final review.

Results

The results are presented on the basis of the five main themes that were constructed during the thematic analysis; namely: (a) *the purpose of group work assessment*, (b) *what is*

assessed in group work, (c) *methods of group work assessment*, (d) *effects and consequences of group work assessment*, and (e) *quality in group work assessment* (see Figure 2). Several subthemes comprise each theme, which in turn mostly includes aspects describing the content of the subtheme. The themes are not mutually exclusive and some aspects occur in more than one theme, alluding to different parts of group work assessment. Furthermore, an aspect may exist within several themes, due to the different content of the articles included in the review. For example, peer assessment reoccurs in several themes but is elucidated from different perspectives. Group work assessment is defined as a comprehensive concept, including all the types of feedback, evaluation, and assessment that are carried out in connection with students' working in groups. Thus, group work assessment includes formative assessment and summative assessment at both individual and group levels. In research, there are also several terms used for grades (e.g., grades and marks); here, the term *grade* is used.

Theme 1: **The Purpose of Group Work Assessment**

The first theme includes research concerning *the purpose of group work assessment*, which encompasses reasons for why assessment in connection with group work is implemented. Assessment can address different purposes. It can involve assessments that are implemented during the ongoing processes to promote students' continued development and learning, which can be addressed as formative assessment. Also, the purpose can be to assess each student's level of knowledge to give grades, which concern summative assessment. This theme comprises the following four subthemes: (a) *improving group work*, (b) *promoting learning*, (c) *giving grades*, and (d) *emulating real life* (Figure 2).

Improving group work. The first subtheme regarding the purpose of group work assessment concern how the assessment can be used to improve the quality of group work by

supporting the group process. It consists of two aspects: (a) *achieving accountability* and (b) *promoting group processes*.

Achieving accountability is about the relationship between group work assessment and accountability (i.e., the degree to which each member of the group participates and does their fair share of the work; Johnson & Johnson, 2004). The importance of creating conditions for group members to take responsibility for the joint assignment is frequently emphasized in the reviewed literature. Using summative assessment (Almond, 2009) at an individual level (Briscoe, 1994) can enhance students' motivation for accepting individual accountability. Using peer assessment can further encourage students to take a more comprehensive perspective on the group work (Kao, 2013).

Group work assessment can be used for the purpose of promoting group processes. Typically occurring problems that may inhibit the group process concern social loafing and free-riding in groups. Group work assessment can be used for the purpose of preventing these phenomena. Several studies have concluded that peer assessment can reduce problems with free-riding in groups (Brooks & Ammons, 2003; Cheng & Warren, 2000; Freeman & McKenzie, 2002), especially if the criteria for assessment are specific and the procedure is pursued multiple times (Brooks & Ammons, 2003). Furthermore, it has been investigated whether diaries both at group and individual levels can reduce social loafing in group work (Dommeyer, 2007). A diary is used as a tool to monitor group members' behavior and should focus on the assignment, problems, and achievements. However, the results of the review reveal that diaries at neither the group nor the individual level reduced social loafing in the groups. Enhancing cooperation by using feedback (Griesbaum & Görtz, 2010) and enhancing teamwork by using in-course assessment (Pratten, Merrick, & Burr, 2014) are other suggested approaches to promote group processes. A further example is to allow students to

evaluate each other's performance and reward individual contributions as a possibility to promote collaboration in groups (Pocock, Sanders, & Bundy, 2016).

Promoting learning. The second subtheme involves approaches in which the purpose of group work assessment is to promote learning. In this sense, formative assessment, or assessment for learning, is used with the intention of giving feedback that supports students' learning process to move forward (McMillan, 2014). Several researchers have concluded that group work assessment has great potential to promote learning (Griesbaum & Görtz, 2010; Hargreaves, 2007; Mutwarasibo, 2013) at both: (a) *the individual level*, and (b) *the group level*.

The individual level concerns the purpose of promoting individual learning in cooperative situations. This can be fulfilled by using feedback approaches that provide insights into how students perceive their individual learning and help teachers to develop learner profiles identifying how each student learns (Griesbaum & Görtz, 2010) or to use web-based peer assessment (Freeman & McKenzie, 2002). Strom and Strom (2011) present a model, The Teamwork Skill Inventory, which is a further example of a method that can be used to identify individual competencies and detect learning needs with the purpose of promoting learning. Assessment at the individual level may also be valuable where feedback is emphasised as important for students' development towards a future profession (Harrison, Lebler, Carey, Hitchcock & O'Bryan, 2013).

The group level concerns how students' engagement in a cooperative learning environment can be used to promote learning. Here, the importance of valid assessment for learning to explore features of cooperative assessment is emphasized (Hargreaves, 2007) and also whether structured cooperative learning activities provide conditions for better learning (Kablan, 2014). When assessing groups with the purpose of promoting learning at a group

level, in-course assessment is presented as one useful option (Pratten et al., 2014) with writing groups as another, both enhancing students' engagement (Mutwarasibo, 2013). Finally, digital technologies are suggested to promote learning in multi-disciplinary teams (Prins, Sluijsmans, Kirschner, & Strijbos, 2005), and facilitate the interaction and cooperation within groups (Webb, 2010).

Giving grades. This subtheme relates to the evaluation of the learning outcome by comparing it against a set of assessment criteria with the purpose of giving grades (McMillan, 2014), and thus refers to a summative assessment. In the reviewed literature, two examples can be found where individual grades are derived from group work. One approach is to give individual grades by deriving them from several assessments (e.g., Pratten et al., 2014; Lejk et. al., 1996; Goldfinch & Raeside, 1990); another approach is to use an individual test of content knowledge within a group project (Pocock et al., 2016). In the latter example, group members cooperated to create a poster and afterwards the group discussed each member's contributions and thereafter allocated grades by following specific criteria.

Emulating real life. This subtheme illustrates how a purpose of using group work assessment can be to emulate real life. Both Alden (2011) and Almond (2009) conclude that the purpose of giving the same grade to a group of students, even if the members may have contributed with different amounts of effort, is that it emulates situations in real life where group members are often rewarded as a group (e.g., groups in a workplace or a sports team). Another argument is that working in groups trains and prepares students with important skills for future group work in professional life (Gammie & Matson, 2007). Group work assessment is seen as necessary for getting students to focus on group work skills, such as the ability to work in groups (Rafiq & Fullerton, 1996). Peer- and self-assessment are especially highlighted as closely related to projects in real life and thus useful if this is the purpose.

Theme 2: What is Assessed in Group Work?

The second theme*, what is assessed in group work?* encompasses the knowledge and skills that are evaluated and measured by the assessment procedure. Research illustrates that students can experience what is to be assessed in group work as vague and often described in general terms by teachers (Forslund Frykedal & Hammar Chiriac, 2011). In some research, a distinction is also made between the assessment of students' different contributions to the group process and the intended outcome of the group's joint work; for instance, learning and/or a product (e.g., Forslund Frykedal & Hammar Chiriac, 2011; Onyia & Allen, 2012; Orr, 2010). This theme comprises the following three subthemes: (a) *contribution*, (b) *knowledge*, and (c) *product* (see Figure 2).

Contribution. The first subtheme concerns the assessment of group members' contributions to group work. Contribution is the most frequent focus of what is assessed in group work in the reviewed research (e.g., Cheng & Warren, 2000; Conway et al., 1993; Goldfinch & Raeside, 1990). *Contribution* comprises three aspects: (a) group interaction (b) intellectual, and (c) workload.

Group interaction includes the different ways in which group members can contribute by interacting with other members during group work: (a) cooperation, (b) positive behaviour, and (c) communication. Cooperation and working together are described by means of several criteria in the reviewed literature, such as cooperative abilities (Forslund Frykedal & Hammar Chiriac, 2011), attendance at meetings or teamwork (e.g., Alden, 2011; Brooks & Ammons, 2003; Strom & Strom, 2011), participating in discussions (Yurdabakan, 2011), helping the group to function well and be efficient as a group (Goldfinch, 2006; Wu et al., 2013), being a good team player (Murray and Boyd, 2015), or contributing with leadership (Wu et al., 2013). Another way for group members to support the interaction

within the group is by contributing with positive behavior; for example, by showing a positive attitude (Lejk et al., 2001a; Onyia & Allen, 2012), being adaptable (Lejk et al., 2001a), showing enthusiasm (Brooks & Ammons, 2003; Goldfinch, 2006), being motivated (Lejk et al., 2001a), getting along with teammates (Strom & Strom, 2011), interacting (Earl, 1986), being sociable (De Wever et al., 2011), or taking responsibility (Yurdabakan, 2011). A third way of contributing to the group interaction is by communicating with teammates (Strom & Strom, 2011), discussing and formulating opinions, by using grounded arguments (De Wever et al., 2011), and negotiating with the group (Alden, 2011), accepting constructive criticism and being a good listener (Lejk & Wyvill, 2001a), as well as seeking and sharing information (Strom & Strom, 2011).

Intellectual contributions, the second aspect, entail different kinds of contribution, such as generating ideas or suggestions (e.g., Butcher et al., 1995; Goldfinch, 2006; Murray & Boyd, 2015), conducting a literature search and analysing the literature (e.g., Cheng & Warren, 2000; Kench et al., 2009; Murray & Boyd, 2015), solving problems (Alden, 2011; Lejk et al., 2001a), thinking critically and creatively (Strom & Strom, 2011), or understanding what is required (Goldfinch, 2006). It can also be about contributing technical skills (Lejk et al., 2001a), or knowing how to use technology properly.

The third aspect concerns the *workload* each group member puts into group work. For instance, contributing with a workload can be about generic contributions, such as getting the job done by: Pulling a fair share (Brooks & Ammons, 2003; Butcher et al., 1995), making an effort (Alden, 2011; Freeman, 1995), or just participating in the group work (Alden, 2011; Kench et al., 2009; Murray & Boyd, 2015). Further examples concern carrying out the task (Butcher et al., 1995), and performing tasks (Alden, 2011; Goldfinch 2006; Yudrabakan, 2011). The contribution may also be more specific, such as participating in meetings (Butcher

et al., 1995; Lejk et al., 2001a), demonstrating effectiveness in meetings (Alden, 2011), keeping group deadlines (Brooks & Ammons, 2003), or contributing to data collection (Johnston & Miles, 2004; Li, 2001).

As seen above, there are various possibilities for group members to contribute to the group's process in different ways. All the criteria found are examples of what in focus in the assessment of contribution. One conclusion that can be drawn from the reviewed literature concerning the subtheme contribution is that it is not really the process that is assessed but instead different forms of contribution to the process of group work. Process is defined as the approach and interaction among students in group work, whereby they contribute in different ways to reaching a joint outcome.

Knowledge. The second subtheme regarding what is assessed in group work encompasses the knowledge that springs from group work. Knowledge is described as an outcome that can be assessed by teachers at both a group and an individual level (van Aalst & Chan, 2007; McKechan & Ellis, 2012) as well as both quantitively and qualitatively in relation to the group's product (McKechan & Ellis, 2012). Theoretical knowledge and cooperative abilities are two aspects of knowledge pointed out by Forslund Frykedal and Hammar Chiriac (2011). The skills observed when students orally present group work are other examples of what is assessed (Earl, 1986; Lejk et al., 2001a). However, given that knowledge is a desired outcome from any process of learning, surprisingly little could be found in the reviewed literature.

Product. The third subtheme regarding what is assessed in group work is product, which is also an outcome that springs from group work (Alden, 2011). Product, along with participation and contribution, are the typical components when grading group work. In the reviewed literature, there are criteria described for assessing the process of work that leads to

the final product. In this sense, process and product become intertwined with each other. One sort of product described is a presentation (e.g., Kench et al., 2009; Li, 2001; Torres Skoumal, 2001). Butcher et al. (1995) and Pocock et al. (2016) give examples in which a poster is assessed at the end of a project; in these cases, both the presentation and each member's contribution were assessed by peers. Other products that are assessed are written reports (e.g., Dingel et al., 2013; Kench et al., 2009; Wu et al., 2013) or conference papers (Wu et al., 2013).

Theme 3: Methods of Group Work Assessment

The third theme concerns *methods of group work assessment*, and comprises the methods that teachers and/or students use to perform different types of group work assessment. In this context, *method* describes how to gather, analyse, and evaluate information about knowledge, skills, products, and processes in group work at both the individual and group level. Methods of assessment include assessment strategies applied from within the group by students through self- and peer-assessment, and from outside the group by the teacher. This theme includes three subthemes: (a) *individual assessment*, (b) *group assessment*, and (c) *combined individual and group assessment* (Figure 2).

Individual assessment. Of the three subthemes regarding methods of group work assessment, methods concerning individual assessment constitute the area that is by far the most thoroughly covered in the reviewed literature. It consists of four aspects: (a) *self-assessment*, (b) *portfolio*, (c) *teacher observation*, and (d) *combinations of peer assessment.*

Self-assessment is a process whereby a student creates a self-report that evaluates their own work (McMillan, 2014). Self-assessment is a useful method for providing information about group processes that are otherwise invisible to teachers (Griesbaum & Görtz, 2010). For example, students' individual logs can be valuable for assessing individual

contributions to group work (Orr, 2010), or to gather information about students' own skills that can be used by a teacher in the process of forming groups (Blowers, 2003).

Portfolio is a collection of documents created by a student and can be used for tracking each member's contribution to the group's work (Alden, 2011). For instance, Singer-Freeman, Bastone, and Skrivanek (2016) used electronic portfolios in a project where students worked in groups but were assessed individually, based on certain criteria for learning. The assessment was not used for grading but only for the students to discuss in their groups.

Teacher observation is one possibility for teachers to perform an informal assessment of students' progress by walking around the classroom to observe the groups working, and thus collecting anecdotal evidence (Gillies & Boyle, 2010). This assessment procedure can also have a formative focus, whereby teachers give instant formative comments on contributions by handing out brief notes to students (McKechan & Ellis, 2012).

Combinations of peer assessment are the most extensively researched aspect regarding methods for individual assessment. Peer assessment can be described as a process whereby students assess each other's learning, contributions and/or efforts (Johnson & Johnson, 2004) and provides a solution for the problem of how to get inside information from the group about their process (Zhang & Ohland, 2009). It is used as a method of extracting information about each individual's contribution to the process within the group work (e.g., Goldfinch & Raeside, 1990; Ko, 2013; Murray & Boyd, 2015). However, to be able to get information about each individual's contribution and calculate individual grades from peer assessment requires methods that adjust the result in order to avoid errors in peer rankings of each other (Bushell, 2006). There are different methods for doing so (i.e., Lejk et al., 1996; Bushell, 2006; Li, 2001). One method is multiplication of the group mark by using an

individual weighting factor (IWF). By using peer assessment of contribution, an IWF can be calculated by multiplying the result of each member's assessment by the group grade. Goldfinch and colleagues (Goldfinch & Raeside, 1990; Goldfinch, 2006) and Conway et al. (1993) used a similar method by allowing students to employ a scheme involving different criteria to assess each other and themselves (see also Johnston & Miles [2004] for a further example). A conclusion from using peer assessment that uses both assessment from peers in a class and peers within the group is that it is perceived as fair by the students (Conway et al., 1993).

According to Onyia (2014), methods using IWF can be categorized by following three possibilities: (a) equal weighting, where peer assessment is worth the same as the teacher's assessment of the final product, (b) combining peer assessment and teachers' assessment from an individual report or presentation; for instance, individual diaries assessed by a teacher can be combined with peer assessment (Rafiq & Fullerton, 1996; Dommeyer, 2007; Keppell et al., 2006) or peer assessment can function alongside an individual essay (Plastow et al., 2010), and (c) unequal weighting, where either the teacher's assessment or the peer assessment is weighted more.

Several researchers have used and developed further methods based on IWF (e.g., Cheng & Warren, 2000; Falchikov, 1993; Wu et al., 2013). Ko (2013) has developed IWF by using an algorithm that is assumed to improve fairness and reliability in the assessment, and that allows teachers to choose whether some criteria should be weighted more than others in the assessment. Li (2000) used an IWF method as a normalisation process whereby grades were adjusted by using a calculation that is supposed to take care of bias and shortcomings in the process of peer assessment, and finally, Murray and Boyd (2015) used the web to pursue

IWF. Results from the reviewed literature indicate that research on the use of IWF to derive individual assessments and grades from peer assessment occurs in a large number of studies.

Another method for deriving an individual grade from peer assessment is to allow peers to distribute a pool of grades among group members by using predetermined categories (e.g., Lejk et al., 1996; Earl, 1986; Freeman, 1995). The results of the peer assessment can then be added to the group grade to generate an individual grade from the teacher. A comparable method allows students to get a plus or minus contribution grade from peer assessment that could help the teacher to adjust individual grades (Lejk et al., 1996), or a reduced individual grade as a penalty for insufficient participation (Kench et al., 2009).

Finally, the following studies which combine different assessment methods (including peer assessment) were found in the review: A computer-based portfolio to capture both collective and individual aspects of learning was studied by Van Aalst (2007). Each portfolio was assessed formatively and summatively by the teacher following set criteria. A combination of several assessments, of which one was peer-assessment, to give an individual grade, was conducted by Edgerton and McKechinie (2002) and Torres-Skoumal (2001). A combination of peer assessment and self-assessment whereby students evaluate their own and peers' skills and attitudes based on set criteria were studied by Strom and Strom (1999), and Strom, Strom, and Moore (1999). The information from these assessments were used to inform the teacher about students' teamwork skills based on the group's interaction. The results of the peer assessment were used both formatively and summatively. Results from the reviewed literature indicate that research combining peer-assessment with other assessment methods at both individual and group level, as well as both formative and summative assessment, occurs in a number of studies.

Group assessment. The second subtheme describes how to assess knowledge, skills, products, and processes at a group level. In the reviewed literature, two aspects of group assessment can be found: (a) *group outcome* and (b) *computer-based group assessment*.

Group outcome encompasses the teacher's assessment of the whole group's work; for instance, when all the members of a group get the same grade based on the group's joint product and effort (Alden, 2011). One method for assessing a group as whole is to let the group take a test together. One example found in the reviewed literature is when students worked in groups with study questions handed out by a teacher. First, the students answered the questions individually, and then gave each other feedback. After the group had studied together, they took a group-based test and received a group grade (Dallmer, 2004).

Group work assessment at the group level can also involve pursuing *computer-based group assessment*. For instance, the group's level of cooperation can be assessed as student work cooperatively on a wiki (van Aals & Chan, 2007; De Wever et al., 2011). This process can be formative as a teacher can use computers to give formative feedback on a presentation to a group (Webb, 2010).

Combined individual and group assessment. The third subtheme concerning group work assessment describes how methods can be combined for both individual and group assessment. In these combinations of methods, the results lead to at least two separate assessments. For instance, if a group task is combined with individual activity, the group task can be assessed at a group level while individual activity is assessed at an individual level (Lejk et al., 1996; Epstein, 2007). One approach is to combine a cooperative written text along with individual journals from each student, which leads to one group assessment alongside individual assessments (Daemmrich, 2010). Another approach found in the reviewed literature is to keep the process of learning in groups separate from the assessment

by using group work only as a way to structure learning. Afterwards, each student turns in an individual paper for assessment, combined with feedback given to the group on their product (Briscoe, 1994). Similar to this approach is another example in which self-assessment was accomplished through individual written learning logs, or group discussion combined with an individual paper assessed by a teacher (Gillies & Boyle, 2010). Another combination of methods is a test that includes two parts, one accomplished individually while the second is solved jointly by working in groups (Crowley, 1997). Finally, there is an example of a method that includes three assessments: two at the group level, consisting of a group assignment followed by group questioning, and the third at the individual level, based on individual questioning (Pratten et al., 2014).

Theme 4: Effects and Consequences of Group Work Assessment

The fourth theme, *effects and consequences of group work assessment*, entails research emphasizing the various effects and consequences arising from using group work assessment. When assessment is applied in group situations, it influences students and teachers in different ways. Using group work assessment causes something to happen, either positive (supporting the affected aspect) or negative (hampering the affected aspect) for the students, either at the individual or group level. Furthermore, the use of group work assessment has consequences for teachers' assessment practice. Hence, this theme comprises three subthemes, two concerning effects on students: (a) *effects on the individual*, (b) *effects on the group*, and the third: (c) *the consequences for teachers' assessment practice* (Figure 2).

Effects on the individual. The first subtheme concerning the effects of group work assessment at the individual level entails three aspects: (a) *learning*, (b) *grades*, and (c) *psycho-social effects*.

Learning concerns the ways in which group work assessment influences an individual's learning. Peer assessment has been pointed out by several researchers as having a positive effect on students' learning. When students assess each group member's answer, it promotes higher achievement compared to teaching methods where students only supply individual answers (Kablan, 2013). Peer assessment focusing on formative assessment with the aim of developing the outcome of the group's work is also considered to have positive effects on students' learning (Keppell et al., 2006). The effect of using peer assessment in combination with self-assessment has been investigated but is not recommended since top-performing students tend to underrate themselves and low-performing students tend to overrate themselves (Lejk et al., 2001b). Still, there are also results from the reviewed research that illustrate positive effects of using self-assessment. For instance, a positive effect on students' academic achievement has been reported (Ali, 2015), especially if feedback is given individually (Archer-Kath, Johnson, & Johnson, 1994). Group work assessment also supports high achievement when using cooperative tests (Hanshaw, 2012) and has a strong positive influence on students' relationship with e-learning (Murray & Boyd, 2015). However, it is the less able students' learning that gains most from group assessment (Ballantine & McCourt Larres, 2007).

The research discussed above does not always explicitly label the type of learning that is the focus of the articles, but discusses learning in general. Some articles concern the learning of skills (i.e., talent, ability, or proficiency in a particular area/competence in doing something) that may come as a result of using group work assessment. Other articles conclude that group work assessment can facilitate students' development of certain abilities, such as generic skills (Ballantine et al., 2007), that self-assessment in group work has a

positive effect on students' self-regulation (Ali, 2015), and promotes students' learning behaviour towards a meta-cognitive orientation (Griesbaum & Görtz, 2010).

The subtheme *grades* concern the effect that group work assessment may have on grades. Research on group grades has revealed that high-achieving students get lower grades in groups than they normally would have received when working alone, and lower-achieving students received higher grades in groups than they normally would have received individually (Moore & Hampton, 2014). Furthermore, Plastow et al. (2010) found that, when students receive group grades, most students get higher grades than they would have got if graded individually.

Psycho-social effects concern the influence that group work assessment can have on a personal level (i.e., the individual's attitude, motivation, stress, and anxiety). If teachers assess individual performance within the group, it can help students to develop a more positive attitude towards group work (Chapman & van Auken, 2001). Students prefer to receive individual grades for their own contribution, and this means that individual assessment and/or structured peer assessment can ease their concerns about grade equity (Orr, 2010). A majority of students also prefer individual assessment over group assessment (Murray & Boyd, 2015) and encounter group assessment with scepticism (Stanier, 1997). Alden (2011) investigated students' preferences regarding group work assessment and found that the methods of recorded review and portfolio are perceived as the fairest and most valid methods for group work assessment, with the least preferred being peer assessment. Students' attitudes towards group assessment show that they find the method unfair and believe that it allows individuals to hide in group work (Lejk et al., 1997). One conclusion is that students prefer types of group work assessment where each group member's contribution is assessed, and these types of group work assessment are also perceived as being most fair and valid.

However, students can also express positive feelings towards the use of group assessment (Ballantine & McCourt, 2007).

Research investigating students' attitudes towards peer assessment reveal that in general they are positive (Mutwarasibo, 2013; Prins et al., 2005). A majority of students perceive peer assessment as fair (Jin, 2012; Onyia & Allen, 2012) and believe that students should play a part in assessment (Jin, 2012; Mutwarasibo, 2013). Nevertheless, students have some concerns regarding the possibility of bias and that they might experience a lack of training in using peer assessment (Walker, 2001). Another concern is that they do not like to assess peers (Mutwarasibo, 2013) because it feels uncomfortable and intimidating (Edgerton & McKechnie, 2002; Goldfinch & Raeside, 1990; Pocock et al., 2016). However, students showed a more positive attitude towards group work and peer assessment after they had participated in the research project (Walker, 2001) or after being given an opportunity to discuss the assessment criteria (Stanier, 1997). Overall, students are more positive towards a holistic approach to peer assessment than a criterion-based one (Lejk et al., 1997).

Group work assessment can also have a positive effect on students' motivation. Giving individual feedback in group work assessment can increase students' achievement and motivation for learning (Archer-Kath et al., 1994). Furthermore, group work assessment can reduce students' levels of anxiety (Daemmrich, 2010; Hanshaw, 2012) and stress (Hanshaw, 2012) when cooperative tests are used, which are perceived positively by both teachers (Hanshaw, 2012) and students (Daemmrich, 2010; Hanshaw, 2012). According to Daemmrich (2010), it is also possible to reduce anxiety over group grades by allowing students to have a second chance by individually redrafting a cooperatively written paper and getting a new grade.

Effects on the group. The second subtheme concerns the effects of group work assessment at the group level. The analysis reveals three aspects connected to group processes and sheds light on whether group work assessment facilitates or hampers the processes of: (a) *cooperation*, (b) *contribution*, and (c) *relations*.

Cooperation encompasses the effects of group work assessment on the cooperation within the group. Results from the reviewed literature indicate that grading seems to be a determining factor. However, the results are not without contradictions. On the one hand, cooperation among group members increases when group grading is used (Stuart, 1994) but, on the other hand, group work assessment can encourage rivalry within groups whereby group work might become a fight for grades among students (Orr, 2010). Another negative aspect of individual grading is that it seems to impair or hinder cooperation, due to reduced pooling and sharing of information between the group members (Hayek et al., 2015). Still, more positive results can be found regarding different types of assessment effects on cooperation. By using peer assessment (De Wever et al., 2011), or self-assessment (van Aals & Chan, 2007), cooperation in the group can be enhanced. Individual feedback can also result in a greater perception of cooperation (Archer-Kath et al., 1994).

Contribution entails research focusing on the effect that group work assessment has on members' contributions when working in cooperative situations. The crucial aspect seems to be type of assessment. Individual grading in group work can influence members' contributions in a positive direction; in fact, when individual assessment was compared to duration of group work and group size, it was the only factor that had any effect on group members' contributions to group work (Joo, 2017). Assessment of individuals and peer assessment (as seen in the previous theme) arises as a possibility, but this can also lead to differentiated contributions (Johnston & Miles, 2004). Students' contributions to the group

can also be influenced by the occurrence of free-riding; i.e., when a student's behaviour (including their contribution) in a group is influenced negatively by the presence of others. Even here, peer assessment is presented as giving the possibility to detect (Griesbaum & Görtz, 2010) and eliminate (Kao, 2013) free-riding in groups by establishing individual accountability and positive interdependence.

Relations focuses on whether relations are influenced by group work assessment. The results of the reviewed literature conclude that group work assessment can actually help building positive relationships among group members by providing individual feedback (Archer-Kath et al., 1994) and build friendships between students by using cooperative tests (Hanshaw, 2012).

Consequences for teachers' assessment practice. The third subtheme about the effects of group work assessment relate to the consequences of the use of group work assessment for teachers' assessment practice. Two aspects have been identified: (a) *grading* and (b) *obtaining information*.

Grading is a central part of assessment. Group work assessment not only influences student grades (see above) but also has consequences for teachers' grading practices. For instance, researchers have examined the effect of using group work assessment in relation to the distribution of grades within groups. Several publications have revealed that peer assessment seems to have the effect of widening the distribution of individual grades given in a group (e.g., Cheng & Warren, 2000; Goldfinch & Raeside, 1990; Zhang & Ohland, 2009).

Obtaining information concerns the challenge facing teachers, when using group work assessment, of obtaining information about work and processes during the group work. It is challenging to assess individuals in a group (Harrison et al., 2013), due to the difficulties teachers experience in acquiring information about each student's contribution to the group

work (Brooks & Ammons, 2003). One reason is that the teacher is not always present while students are working in groups, which consequently leaves the teacher with a lack of information to apply to the assessment procedure (Murray & Boyd, 2015). Leaving teachers with a lack of information is troubling since it is hard to determine each student's contribution to the work (McKechan & Ellis, 2012).

Theme 5: Quality in Group Work Assessment

The fifth theme concerns the criteria used to evaluate *quality in group work assessment*. These quality criteria rate whether the group work assessment method being used is reliable and fair, from both the students' and the teachers' perspectives. Validity, reliability, and fairness are the criteria applied in the evaluation of quality (McMillan, 2014). Validity concerns how well the assessment corresponds to the aim of the group work, while reliability is about the precision of the assessment. Fairness is about how to provide equal conditions for all students in the group and how to perform the assessment without bias. Validity and reliability are integrated within several references and therefore difficult to separate. Nevertheless, regarding validity, it is argued that a valid assessment should support its purpose (Hargreaves, 2007). Concerning learning in a cooperative environment, a valid assessment should support both learning and group processes. Results from the reviewed literature also indicate that group work assessment has the potential to improve validity if skills and competencies related to real-life situations are included in the assessment (see the aspect *emulating real life*, Gammie & Matson, 2007).

Group work assessment is a multidimensional concept which includes the use of a number of methods of assessment in conjunction with group work. Consequently, one cannot talk about quality in group work assessment as though there were only *one* method, but must refer to the method mentioned. Therefore, this theme has the following subthemes: (a) *quality*

in peer assessment, (b) *quality in methods for individual assessment*, (c) *quality in group grading*, and (d) *quality in method combinations* (Figure 2).

Quality in peer assessment. The first subtheme in relation to quality of group work assessment concerns one of the most frequently used methods in group work assessment, namely, peer assessment. In previous themes when reviewing methods we have concluded that peer assessment is often used in combination with individual assessment. Since there is so much research concerning peer assessment, this constitutes a subtheme on its own, encompassing two aspects: (a) *validity and reliability* and (b) *fairness.*

Validity and reliability. The review reveals an ambiguous result in terms of the reliability of peer assessment; some results speak of high reliability (Johnston & Miles, 2004; De Wever et al., 2011), but a larger number speak of low validity and reliability, implying that using peer assessment can reduce the validity and reliability of group work assessment (e.g., Gammie & Matson, 2007; Murray & Boyd, 2015; Orr, 2010). Problems pertaining to reliability in connection with peer assessment include a low correlation between peer evaluations and students' course performance (Dingel et al., 2013), or when the results of peer assessment are compared with self-assessment (Johnston & Miles, 2004), or with whole-course assessment (Dingel et al., 2013). Different interpretations of grading scales, errors in intra-group rankings, or problems with free-riders also affect the reliability negatively (Johnston & Miles, 2004). Reliability in terms of comparing teachers' assessments with peer assessment show a similar pattern. According to Freeman (1995), there is no difference in average scores when comparing peer and teacher assessment. Other researchers highlight problems in reliability related to differences in what is assessed (i.e. peers tend to assess effort instead of skill; Dingel et al., 2013), and rating of performance (i.e. peers tend to overrate poor performances and underrate good performances; Freeman, 1995). According to

Cheng and Warren (1999), peer assessment is not sufficiently reliable, although it is possible to reduce the differences between peer and teacher assessments by practice in assessment.

When it comes to the students' perspective on the reliability of peer and self-assessment, they have little faith in these methods (Griesbaum & Görtz, 2010). Lejk and Wyvill (2001b) suggest that teachers refrain from using self-assessment as a summative assessment, since there is a tendency for top-performing students to underestimate their contribution and vice versa. Students raise even more doubts about peer summative assessment, which they regard as a flawed and unserious practice. Still, reliability and validity can be improved if students are assessed individually as well. But there are also results that are in opposition to these findings. Most of the methods presented that weight individual contributions by using self- and peer-assessment are also unreliable (Spatar et al., 2015) and it is hard to interpret them correctly when deriving individual grades from group performances (Almond, 2009), which may cause problems with reliability using this kind of technique (Kablan, 2014). Furthermore, Mathew (1993) concluded that peer assessment is not a robust enough assessment strategy to be used as a basis for distributing grades.

Fairness in peer assessment is one factor that may influence the quality of assessment. Peer assessment can be one strategy used to increase fairness in group work assessment (e.g., Gammie & Matson, 2007; Ko, 2013; Onyia, 2014) because it can make each individual effort more apparent. To improve and structure a fair peer assessment, some prerequisites are proposed; that students: (a) pursue the assessment of each other's contributions progressively, with concrete documents and evidence (Onyia & Allen, 2012), (b) use set criteria to structure the assessment, including an individual weighting factor strategy in the form of an algorithm (Ko, 2013), and (c) use anonymous peer assessment (Lejk et al., 2001b; Freeman & McKenzie, 2002; Kench et al., 2009). Fairness in peer

assessment can also be improved by combining it with a teacher assessment in which each individual contribution is calculated (Cheng & Warren, 2000). However, results from the reviewed research also elucidate that peer assessment may decrease fairness. One of the most frequently occurring worries that students express concerns bias (Walker, 2001). There is research concerning bias in peer assessment compared with self-assessment (Johnston & Miles, 2004). One conclusion is that students rate their own contribution higher than those of other group members. Since self- and peer-assessment is a subjective process, the factor of bias has to be accounted for when extracting individual grades from group grades. One way to handle the problem of bias in peer assessment could be to use a method that includes normalization (Li, 2001), although this idea has been criticized by Bushell (2006), who instead suggests a method whereby the rank order between peer assessors should be preserved as far as possible.

Different types of bias may occur in peer assessment in group work. In the reviewed literature, we found that bias can be based on: (a) friendship, (b) gender, (c) race, and (d) group role. Bias in peer assessment based on friendships and social interactions is described as a factor that may influence the results of peer assessment (Magin, 2001). Some researchers point out from their own experience with peer assessment that students find it hard to be unbiased and to remember everything that happened in the group during the course (Goldfinch & Raeside, 1990). Bias seems to occur unconsciously, but one possibility to reduce it is to use methods that correct for errors in peer assessment (Bushell, 2006). Further results also indicate that friends tend to give higher scores to each other (Kench et al., 2009; Falchikov, 1986). However, Magin (2001) investigated the bias effect of friendship and social interaction and found that it had very little impact, which implies that it is possible to

perform peer assessment relatively free from bias if the students fully understand the terms of such assessment.

Bias can also be based on gender (Falchikov & Magin, 1997). In one study, the difference between boys' and girls' scores was investigated (Yurdabakan, 2011). Girls came closer to the teachers' score than boys when assessing each other's contributions. Boys assessed both other boys and girls equally but girls on the other hand gave other girls and boys lower scores in comparison to the boys.

Finally, race and group role are additional factors that can introduce bias when using peer assessment (Dingel & Wei, 2013). For example, white students and leaders seem to receive higher evaluations from their peers than others (Dingel & Wei, 2013).

Quality in methods for individual assessment. Even if peer assessment, problematized in the previous subtheme, can be considered as a method for individual assessment, there are also other methods for this. Thus, this subtheme relates to the quality of other methods for individual assessment. Two aspects can be found: (a) *validity and reliability* and (b) *fairness.*

Validity and reliability. One aspect that makes individual assessment unreliable is the fact that it is challenging for the teacher to assess individual contributions to a task that has been carried out within a cooperative setting (Gillies & Boyle, 2010; Harrison et al., 2013; Orr, 2010). This is related to the fact that teachers do not have the opportunity to always be present during the group's work (Murray & Boyd, 2015).

Individual assessment in group work can also be carried out by using portfolios. Singer-Freeman et al. (2016) investigated the reliability of using electronic portfolios for assessing individuals within group work. The conclusion drawn from their study is that this is a reliable way to document cooperative learning. However, the study also investigated

whether there are any differences between assessments by teachers who know the students and teachers who do not. The conclusion was that the teachers who knew the students found more evidence for learning than the teachers who did not know them. This might imply some problems with reliability when using portfolios.

Fairness. Individual grading has been emphasized in several studies as a way to improve fairness in group work assessment (Gammie & Matson, 2007; Ko, 2013; Onyia, 2014). Students state that assessment in group work is unfair (Harrison et al., 2013) if equal grades are given to all group members for unequal contributions (Conway et al., 1993; Freeman & McKenzie, 2002). One approach to improving fairness in individual assessments of group work is by using wikis (Benckendorff, 2009), since the assessment of each individual's product and the grading of an individual within the group becomes more transparent. Wikis also reduce student concerns about the distribution of grades because each individual's efforts within the group become more apparent (Caple & Bogle, 2011).

There are also factors that decrease fairness in individual assessment. For example, it is a challenge for teachers to be able to assess each individual contribution to group work and make a fair assessment (Harrison et al., 2013; Orr, 2010). Teachers' lack of experience regarding group work assessment methods (Strom & Strom, 2011), especially the lack of individual assessment methods (Joo, 2017), could also negatively influence fairness and hamper teachers' performance of group work assessment (Strom & Strom, 2011).

Quality within group grading. The third subtheme concerns quality within group grading (i.e. when a group is rewarded with a joint grade). No significant correlation between individual grades and group grades (Plastow et al., 2010), or between group assessment and individual assessment (Epstein, 2007) have been found. Students who fail an individual assignment can pass the group assignment, and vice versa. Consequently, one conclusion that

32

can be drawn is that a group grade is not an accurate indicator of students' individual performance (Plastow et al., 2010). Additionally, the validity of different methods for deriving an individual grade from a group grade by combining a group grade with peer assessment of contributions has been investigated (Zhang & Ohland, 2009). Three conclusions were drawn from this study. First, a group grade without any adjustments made is not a valid way to measure students' abilities. Second, group size affects the validity of a group grade, with the reliability decreasing in larger groups. Third, the study concludes that the validity is improved by using a method of peer assessment of contributions combined with group grading. Finally, according to students, group grading is a method that does not reflect the individual contributions to a group work, which gives additional support to the claim that this method may be both unreliable and invalid (Strauss & U, 2007).

Quality in method combinations. The quality of group work assessment can be influenced by using a combination of both individual and group assessment. However, combining two or three methods does not enhance the quality of group work assessment per se (Edgerton & McKechinie, 2002). On the contrary, these authors show that a combination of peer assessment, group assessment, and individual questioning does not correlate with the students' performance in an individual exam. This result is supported by Almond (2009), who shows that a combination of group assessment and individual assessment, in which individual grades were extracted from teacher, peer- and self-assessments and then compared with a summative group assessment, did not enhance the quality. The results revealed that students with high individual grades got lower grades in the group assessment and vice versa when they were assessed individually. A conclusion that can be drawn from these two publications is that the use of several assessment methods does not obviously enhance the quality of group work assessment.

Discussion

　　Based on the results of our review, we have chosen to critically discuss and problematize what we consider to be the most important findings.

Group Work Assessment Becomes the Assessment of Social Skills

　　One key finding of this review is that, for the most part, group work assessment turns into an assessment of social skills focusing on group processes (e.g., Griesbaum & Görtz, 2010; Pratten, et al., 2014). Primarily contribution, but also performance, participation, and engagement often recur as the focus in group work assessment, while very little research addresses the assessment of knowledge acquisition. In fact, in the reviewed research, we found no articles discussing or problematizing the absence of knowledge assessment as a learning objective. Our inference is that teachers use group work assessment as a means of structuring cooperation and making sure that each member contributes to the group's cooperative work. Hence, instead of assessing knowledge and ensuring that the students achieve their learning goals, teachers emphasize evaluation of the extent to which each group member participates and contributes to the group work. In this sense, group work assessment functions as a means of structuring the learning of group work skills and not as an objective for assessing knowledge. We argue, in line with McMillan (2014), that in classroom assessment the goals, objectives, and standards for assessment are basic and important tools for ensuring reliable and fair assessment. Furthermore, in many countries, including the USA, the UK, and Sweden, according to the curriculum, assessment should be based on goals that are specified in terms of individual knowledge (Lundahl, Hultén, & Tveit, 2016). The teachers' tasks of assessing students' learning and knowledge are prescribed in the curriculum, and it is therefore both surprising and remarkable that so little research on group

work assessment takes the teachers' prescribed assignments into consideration. This area therefore needs more attention in further research.

Focus on Peer Assessment

Another crucial insight from this review is that a great deal of the existing research focuses on peer assessment (e.g., Wu et al., 2013; Ko, 2013; Murray & Boyd, 2015). Most of the research using peer assessment as a method of group work assessment employs it to resolve the challenge of assessing individuals within a group. However, it might not be surprising that peer assessment is so frequently used in group work assessment, since one problem emphasized in the reviewed research on group work assessment is to identify *who did what in the group*; that is, each individual's contribution and/or learning (i.e., Ko, 2013; Murray & Boyd, 2015). Here, peer assessment emerges as a solution for gleaning inside information about the group's process and work from the insider perspective of the members themselves. Based on the assumption that the students have taken part in the group work, they can contribute with unique information about each group member as well as the joint group work (e.g., Ko, 2013; Murray & Boyd, 2015; Alden, 2011).

A conclusion of this review is that quality in group work assessment is an area that has generated a lot of studies. We argue that quality in group work assessment must be specified by defining the quality aspects of the different methods that constitute the large and complex concept of group work assessment. When reviewing the quality of different methods of group work assessment, peer assessment emerges as the most extensive subtheme. However, this review also reveals that there might be problems with the quality of peer assessment in terms of validity and reliability (Dingel et al., 2013; Lejk & Wyvill, 2001b; Kablan, 2014), and that there is often a difference between peer and teacher assessment (Cheng & Warren, 1999). Consequently, there is a need for research focusing on reliability,

validity, and fairness in conjunction with investigations into why peer assessment is so highly prioritized and extensive, considering that there is to date little evidence that it can provide high-quality assessment. We argue that even more research needs to be carried out concerning the assessment of each individual's achievements during group work in order to gain more knowledge about how to implement more high-quality group work assessment for each individual student.

Absence of the Teacher in Individual Assessment within the Group

A third insight drawn from this review is that there is a lack of research investigating the teacher's role in group work assessment. By allowing peers to assess each other, it seems as though the teacher is stepping aside and is absent from the group work assessment process, thus leaving the problem in the hands of the students to solve. But when teachers delegate their assessment assignment to students, they entrust them with a difficult and complex task. Is this reasonable when group work assessment seems to be an overwhelming task, even for educated and experienced teachers (e.g. Forslund Frykedal & Hammar Chiriac, 2011, 2016; Murray & Boyd, 2015)? This review also reveals that peer assessment is often used in combination with the teacher's assessment. The role of peer assessment has been extensively explored but how the teacher's assessment is pursued has not been investigated in much detail. This gap reveals a need for future research focusing on investigating the teacher's role and assignment in group work assessment.

Conclusions

In this review of existing research on group work assessment in educational settings, an integrated research area consisting of research combining group work and classroom assessment has been systematized and developed into a typology presenting an overview of the research field. Additionally, through this review, some important findings have been

identified, giving us new insights and ideas for future research. The focus on: (a) social skills and group processes as an objective for group work assessment, (b) peer assessment and quality in conjunction with peer assessment, and (c) the absence of teachers from individual assessment in group work assessment, disclose gaps in the research area that need further attention. Future research on: (a) how to assess knowledge as a learning objective, (b) how to improve quality in group work assessment by considering the quality aspects of reliability, validity, and fairness and, finally, (c) the teacher's role and assignment in group work assessment, are essential questions for future research. In sum, this review adds new insights and brings a scientific contribution to the research field of group work assessment.

Limitations

During the process of writing this review, some limitations have come to our attention. Firstly, the range of the review, including 83 articles and dissertations, crafts a *result-heavy* review. A more stringent selection, for instance, focusing on individual assessment in group work assessment, could have gained in both depth and scope, providing more space for a critical discussion. Secondly, there is no distinction made between different school contexts, which might have led to a rather sprawling review. For instance, there might be differences in the use of methods and the focus on what is assessed in different school contexts, such as higher education compared with compulsory school. Thirdly, there are many different terms figuring in the research field of group work assessment. Some terms describe the same phenomenon but in different words. In the process of categorizing and analysing, it was necessary to interpret these terms, which might have caused some misinterpretations.

Ethical considerations

The study was approved by the regional Research and Ethics Committee at Linköping University, Sweden (Dnr 2013/401-31, Dnr 2014/134-32 and Dnr 2016/209-32).

Funding

Disclosure statement

The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

*Alden, J. (2011). Assessment of individual student performance in online team projects. *Journal of Asynchronous Learning Network*, *15*, 5–20. doi:10.24059/olj.v15i3.193

*Ali, H. (2015). The effect of collaborative learning and self-assessment on self-regulation. *Educational Research and Reviews*, *10*, 2164–2167. doi:10.5897/ERR2015.2349

Alm, F., & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment*, *20*, 132–150. doi:10.1080/10627197.2015.1028620

*Almond, R. J. (2009). Group assessment: Comparing group and individual undergraduate module marks. *Assessment & Evaluation in Higher Education*, *34*, 141–148. doi:10.1080/02602930801956083

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, *27*(2), 3–13. doi:10.1111/j.1745-3992.2008.00118.x

*Archer-Kath, J., Johnson, D. W., & Johnson, R. T. (1994). Individual versus group feedback in cooperative groups. *The Journal of Social Psychology*, *5*, 681–694.

Baines, E., Blatchford, P., & Chowne, A. (2007). Improving the effectiveness of collaborative group work in primary schools: Effects on science attainment. *British Educational Research Journal*, *33*, 663–680. doi:10.1080/01411920701582231

*Ballantine, J., & Mccourt Larres, P. (2007). Final year accounting undergraduates' attitudes to group assessment and the role of learning logs. *Accounting Education*, *16*, 163–183. doi:10.1080/09639280701234419

*Benckendorff, P. (2009). Evaluating wikis as an assessment tool for developing collaboration and knowledge management skills. *Journal of Hospitality and Tourism*

Management, *16*, 102–112. doi:10.1375/jhtm.16.1.102

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers'
summative assessments. *Assessment in Education: Principles, Policy and Practice*, *17*,
215–232. doi:10.1080/09695941003696016

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning:
Putting into practice.* Buckingham, UK: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in
education: Principles, policy & practice 5*(1), 7–74. doi:10.1080/0969595980050102

*Blowers, P. (2003). Using student skill self-assessments to get balanced group for group
project. *College Teaching*, *51*, 106-110.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research
in Psychology*, *3*(2), 77–101. doi:10.1191/1478088706qp063oa

*Briscoe, C. (1994). Making the grade: Perspectives on a teacher's assessment practices.
Mid-Western Educational Researcher, *7*(4), 14-16.

*Brooks, C. M., & Ammons, J. L. (2003). Free riding in group projects and the effects of
timing, frequency, and specificity of criteria in peer assessments. *Journal of Education
for Business*, *78*, 268–272. doi:10.1080/08832320309598613

*Bushell, G. (2006). Moderation of peer assessment in group projects. *Assessment &
Evaluation in Higher Education*, *31*, 91–108. doi:10.1080/02602930500262395

*Butcher, A. C., Stefani, L. A. J., & Tariq, V. N. (1995). Analysis of peer-, self- and staff-
assessment in group project work. *Assessment in Education: Principles, Policy &
Practice*, *2*, 165–185. doi:10.1080/0969594950020204

*Caple, H., & Bogle, M. (2011). Making group assessment transparent: What wikis can
contribute to collaborative projects. *Assessment & Evaluation in Higher Education*, *38*,

198–210. doi:10.1080/02602938.2011.618879

*Chapman, K. J., & van Auken, S. (2001). Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education*, *23*, 117–127. doi:10.1177/0273475301232005

*Cheng, W., & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment & Evaluation in Higher Education*, *24*, 301-314. doi:10.1086/250095

*Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in Higher Education*, *5*, 243–255. doi:10.1080/135625100114885

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

*Crowley, M. (1997). Aligning assessment with classroom practices: A promising testing format. *The Mathematics Teacher*, 706–711. doi:10.1108/17506200710779521

*Daemmrich, I. G. (2010). Assessing collaborative writing in nontraditional and traditional first-year college writing courses. *Teaching English in the Two-Year College*, *38*, 161–175.

*Dallmer, D. (2004). Collaborative test taking with adult learners. *Adult Learning*, *15*(3-4), 4–8. doi:org/10.1177/104515950401500301

Davidson, N., & Major, C. H. (2014). Boundary crossings: Cooperative learning, collaborative learning, and problem-based learning. *Journal on Excellence in College Teaching*, *25*(3-4), 7–55.

*De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in

a wiki: The reliability of university students' peer assessment. *Internet and Higher Education*, *14*, 201–206. doi:10.1016/j.iheduc.2011.07.003

*Dingel, M. J., Wei, W., & Huq, A. (2013). Cooperative learning and peer evaluation: The effect of free riders on team performance and the relationship between course performance and peer evaluation. *Journal of the Scholarship of Teaching and Learning*, *13*(1), 45–56.

*Dingel, M., & Wei, W. (2013). Influences on peer evaluation in a group project: An exploration of leadership, demographics and course performance. *Assessment & Evaluation in Higher Education*, *39*, 729–742. doi:10.1080/02602938.2013.867477

*Dommeyer, C. J. (2007). Using the diary method to deal with social loafers on the group project: Its effects on peer evaluations, group behavior, and attitudes. *Journal of Marketing Education*, *29*, 175–188. doi:10.1177/0273475307302019

*Earl, S. E. (1986). Staff and peer assessment: Measuring an individual's contribution to group performance. *Assessment & Evaluation in Higher Education*, *11*, 60–69. doi:10.1080/0260293860110105

*Edgerton, E., & McKechnie, J. (2002). Students' views of group-based work and the issue of peer assessment. *Psychological Bulletin*, *2*(2), 76–81.

*Epstein, E. (2007). *Blending individual and group assessment: A model for measuring student performance using mathematics as an example*. (Doctoral dissertation, Columbia University). Available from ProQuest Dissertations and Theses database. (UMI No. 3259244)

*Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment & Evaluation in Higher Education*, *11*, 146–166. doi:10.1080/0260293860110206

*Falchikov, N. (1993). Group process analysis: Self and peer assessment of working together in a group. *Education and Training Technology International*, *30*, 275–284.

*Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment and Evaluation in Higher Education*, *22*, 385–396. doi:10.1080/0260293970220403

Forslund Frykedal, K. (2008). *Elevers tillvägagångssätt vid grupparbete. Om amtionsnivå och interaktionsmönster i samarbetssituationer. [Students procedure at group work. About ambitions and interaction in cooperative situations]* Linköping University. (Doctoral dissertation, Linköping University)

*Forslund Frykedal, K., & Hammar Chiriac, E. (2011). Assessment of students' learning when working in groups. *Educational Research*, *53*, 331–345. doi:10.1080/00131881.2011.598661

Forslund Frykedal, K., & Hammar Chiriac (2016). To make the unknown known: Assessment in group work among students. *Journal of Educational Research, 2,* 149-162.

Forslund Frykedal, K., & Hammar Chiriac, E. (2018). Student collaboration in group work: Inclusion as participation. *International Journal of Disability, Development and Education*, *65*, 183–198. doi:10.1080/1034912X.2017.1363381

*Freeman, M. (1995). Peer assessments by groups of group work. *Assessment & Evaluation in Higher Education*, *20,* 289–300.

*Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student teamwork: Benefits of evaluating across different subjects. *British Journal of Educational Technology*, *33*, 551–569.

*Gammie, E., & Matson, M. (2007). Group assessment at final degree level: An evaluation.

Accounting Education, *16*, 185–206. doi:10.1080/09639280701234609

Gillies, R. M. (2007). *Cooperative learning: Integrating theory and practice*. Thousand

Oaks, CA: Sage.

Gillies, R. M. (2017). Promoting academically productive student dialogue during

collaborative learning. *International Journal of Educational Research*.

doi:10.1016/j.ijer.2017.07.014

*Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of

implementation. *Teaching and Teacher Education*, *26*, 933–940.

doi:10.1016/j.tate.2009.10.034

Gillies, R. M., & Boyle, M. (2011). Teachers' reflections on cooperative learning (CL): A

two-year follow-up. *Teaching Education*, *22*, 63–78. doi:10.1080/

10476210.2010.538045

Gillies, R. M., & Khan, A. (2008). The effects of teacher discourse on students' discourse,

problem-solving and reasoning during cooperative learning. *International Journal of

Educational Research*, *47*, 323–340. doi:10.1016/j.ijer.2008.06.001

*Goldfinch, J., & Raeside, R. (1990). Development of a peer assessment technique for

obtaining individual marks on a group project. *Assessment and Evaluation in Higher

Education*, *15*, 210–231.

*Goldfinch, J. (2006). Further developments in peer assessment of group projects.

Assessment & Evaluation in Higher Education, *19*, 29–35.

doi:10.1080/0260293940190103

*Griesbaum, J., & Görtz, M. (2010). Using feedback to enhance collaborative learning: An

exploratory study concerning the added value of self- and peer-assessment by first-year

students in a blended learning lecture. *International Journal of E-Learning and Distance*

Education, *9*, 481–503.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*, 309–333. doi:10.1207/S15324818AME1503_5

Hammar Chiriac, E. (2003). *Grupprocesser i utbildning [Group processes in education]*. Linköping, Sweden: Linköping University. (Doctoral dissertation, Linköping University)

Hammar Chiriac, E. (2008). A scheme for understanding group processes in problem-based learning. *Higher Education*, *55*, 505–518. doi:10.1007/s10734-007-9071-7

Hammar Chiriac, E. (2014). Group work as an incentive for learning: Students' experiences of group work. *Frontiers in Psychology*, *5*(June), 1–10. doi:10.3389/fpsyg.2014.00558

Hammar Chiriac, E., & Granström, K. (2012). Teachers' leadership and students' experience of group work. *Teachers and Teaching: Theory and Practice*, *18*, 345–363. doi:10.1080/13540602.2012.629842

*Hanshaw, L. G. (2012). Qualitative aspects of group-only testing. *College Student Journal*, *46*, 419–426.

*Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *14*, 185–199. doi:10.1080/09695940701478594

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, *20*, 245–270. doi:10.1080/02671520500193744

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles,*

Policy & Practice, *4*, 365–379. doi:10.1080/0969594970040304

*Harrison, S. D., Lebler, D., Carey, G., Hitchcock, M., & O'Bryan, J. (2013). Making music

or gaining grades? Assessment practices in tertiary music ensembles. *British Journal of*

Music Education, *30*, 27–42. doi:10.1017/S0265051712000253

*Hayek, A. S., Toma, C., Oberlé, D., & Butera, F. (2015). Grading hampers cooperative

information sharing in group problem solving. *Social Psychology*, *46*(3), 121–131.

doi:10.1027/1864-9335/a000232

Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: A

comparison. *Curriculum Journal*, *16*, 153–176. doi:10.1080/09585170500135954

*Jin, X. (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*, 577–589.

doi:10.1080/02602938.2011.557147

Johnson, D. W., & Johnson, R. T. (2004). *Assessing students in groups: Promoting group*

responsibility and individual accountability. Thousand Oaks, CA: Corwin Press.

*Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment*

& Evaluation in Higher Education, *29*, 751–768. doi:10.1080/0260293042000227272

*Joo, M.-H. (2017). Students' group work contribution: Influence of work preference,

gender, and individual assessment. *Social Behavior and Personality*, *45*, 19–28.

doi:10.2224/sbp.5385

*Kablan, Z. (2014). Comparison of individual answer and group answer with and without

structured peer assessment. *Research in Science & Technological Education*, *32*, 251–

262. doi:10.1080/02635143.2014.931840

Kamps, D., Wendland, M., Culpepper, M., Lamb, D., & Charter, E. (2006). Active teacher

participation in functional behavior assessment for students with emotional and behavioral disorders risks in general education classrooms. *Behavioral Disorders*, *31*, 128–146.

*Kao, G. Y. M. (2013). Enhancing the quality of peer review by reducing student "free riding": Peer assessment with positive interdependence. *British Journal of Educational Technology*, *44*, 12–124. doi:10.1111/j.1467-8535.2011.01278.x

*Kench, P. L., Field, N., Agudera, M., & Gill, M. (2009). Peer assessment of individual contributions to a group project: Student perceptions. *Radiography, 15*, 158–165. doi:10.1016/j.radi.2008.04.004

*Keppell, M., Au, E., Ma, A., & Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment & Evaluation in Higher Education*, *31*, 453–464. doi:10.1080/02602930600679159

*Ko, S. (2013). Peer assessment in group projects accounting for assessor reliability by an iterative method. *Teaching in Higher Education*, *19*, 301–314. doi:10.1080/13562517.2013.860110

*Lejk, M., & Wyvill, M. (2001a). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, *26*, 61–72. doi:10.1080/0260293022000020327

*Lejk, M., & Wyvill, M. (2001b). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education, 26* 551–561. doi:10.1080/02602930120093887

*Lejk, M., Wyvill, M., & Farrow, S. (1996). A survey of methods of deriving individual grades from group assessments. *Assessment & Evaluation in Higher Education*, *21*, 267-

280.

*Lejk, M., Wyvill, M., & Farrow, S. (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*, 81–91.

doi:10.1080/0260293970220106

*Li, L. K. Y. (2001). Some refinements on peer assessment of group projects. *Assessment &*

Evaluation in Higher Education, *26*(1), 5-18. doi:10.1080/026029300200225

Lundahl, C., Hultén, M., & Tveit, S. (2016). *Betygssystem i internationell belysning*

[Grading systems in a international light]. Stockholm, Sweden: Skolverket.

*Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work

reciprocity as a source of bias in multiple peer assessment of group work. *Studies in*

Higher Education, *26*(1), 53-63. doi:10.1080/0307507002003071

*Mathews, B. (1993). Assessing individual contributions: Experience of peer evaluation in

major group projects. *British Journal of Educational Technology*, *25*, 19-28.

*McKechan, S., & Ellis, J. (2012). Collaborative learning in the Scottish curriculum for

excellence: The challenges of assessment and potential of multi-touch technology.

Education 3–13, *42*, 475–487. doi:10.1080/03004279.2012.717959

McMillan, J. H. (Ed.). (2012). *SAGE handbook of research on classroom assessment*.

Thousand Oaks, CA: Sage.

McMillan, J. H. (2014). *Classroom assessment: principles and practice for effective*

standard-based instruction. New York, NY:Pearson.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Academia and clinic annals of

internal medicine preferred reporting items for systematic reviews and meta-analyses:

The PRISMA statement. *Annulas of Internal Medicine*, *151*, 264–269.

doi:10.1371/journal.pmed1000097

*Moore, P., & Hampton, G. (2014). "It's a bit of a generalisation, but…": Participant

perspectives on intercultural group assessment in higher education. *Assessment &*

Evaluation in Higher Education, *40*, 390–406. doi:10.1080/02602938.2014.919437

*Murray, J. A., & Boyd, S. (2015). A preliminary evaluation of using WebPA for online peer

assessment of collaborative performance by groups of online distance learners.

International Journal of E-Learning and Distance Education, *30*, 112-124.

*Mutwarasibo, F. (2013). *Understanding group-based learning in an academic context:*

Rwandan students' reflections on collaborative writing and peer assessment. Linköping,

Sweden: Linköping University. (Doctoral dissertation, Linköping University)

*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.

*Onyia, O. P., & Allen, S. (2012). Peer assessments of GPW: Infusing fairness into students'

assessments of peer contributions. *Research in Higher Education Journal*, 1–20.

*Orr, S. (2010). Collaborating or fighting for the marks? Students' experiences of group

work assessment in the creative arts. *Assessment & Evaluation in Higher Education*, *35*,

301–313. doi:10.1080/02602931003632357

*Plastow, N., Spiliotopoulou, G., & Prior, S. (2010). Group assessment at first year and final

degree level: A comparative evaluation. *Innovations in Education and Teaching*

International, *47*, 393–403. doi:10.1080/14703297.2010.518427

*Pocock, T. M., Sanders, T., & Bundy, C. (2016). The impact of teamwork in peer

assessment : A qualitative analysis of a group exercise at a UK medical school.

Bioscience Education, *15*, 1–12. doi:10.3108/beej.15.3

*Pratten, M. K., Merrick, D., & Burr, S. A. (2014). Group in-course assessment promotes cooperative learning and increases performance. *Anatomical Sciences Education*, *7*), 224–233. doi:10.1002/ase.1397

*Prins, F. J., Sluijsmans, D. M., Kirschner, P. A., & Strijbos, J. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment & Evaluation in Higher Education*, *30*, 417–444. doi:10.1080/02602930500099219

*Rafiq, Y., & Fullerton, H. (1996). Peer assessment of group projects in civil engineering. *Assessment & Evaluation in Higher Education*, *21*, 69–81. doi:10.1080/0260293960210106

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues & Practice*, *24*(2), 3–13. doi:10.1111/j.1745-3992.2005.00006.x

*Singer-freeman, K., Bastone, L., & Skrivanek, J. (2016). Using eportfolios to assess applied and collaborative learning and academic identity in a summer research program for community college students. *International Journal of ePortfolio*, *6*, 45–57.

*Spatar, C., Penna, N., Mills, H., Kutija, V., & Cooke, M. (2015). A robust approach for mapping group marks to individual marks using peer assessment. *Assessment & Evaluation in Higher Education*, *40*, 371–389. doi:10.1080/02602938.2014.917270

*Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education*, *21*, 95–98. doi:10.1080/03098269708725413

*Strauss, P., & U, A. (2007). Group assessments: Dilemmas facing lecturers in multicultural tertiary classrooms. *Higher Education Research & Development*, *26*, 147–161.

doi:10.1080/07294360701310789

*Strom, P. S., & Strom, R. D. (2011). Teamwork skills assessment for cooperative learning. *Educational Research and Evaluation*, *17*, 233–251. doi:10.1080/13803611.2011.620345

*Strom, P. S., Strom, R. D., & Moore, E. G. (1999). Peer and self-evaluation of teamwork skills. *Journal of Adolescence*, *22*, 539–553. doi:10.1006/jado.1999.0247

*Stuart, M. A. (1994). Effects of group grading on cooperation and achievement in two fourth-grade math classes. *The Elementary School Journal*, *95*, 11–21.

*Torres-Skoumal, M. (2001). Alternative assessment models: Assessment through group work and the use of CAS as a self-assessment tool. *International Journal of Computer Algebra in Mathematics Education*, *8*, 61-84.

van Aalst, J. (2013). Assessment in collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 280–296). New York, NY: Routledge.

*van Aalst, J., & Chan, C. K. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *Journal of the Learning Sciences. 16*, 175–220. doi:10.1080/10508400701193697

*Walker, A. (2001). British psychology students' perceptions of group-work and peer assessment. *Psychology Learning & Teaching*, *1*, 28–36. doi:10.2304/plat.2001.1.1.28

*Webb, M. (2010). Beginning teacher education and collaborative formative e-assessment. *Assessment & Evaluation in Higher Education*, *35*, 597–618. doi:10.1080/02602931003782541

Webb, N., Franke, M., Ing, M., Turrou, A., & Johnson, N. (2015). Student participation, teacher instructional practices, and the developmental of mathematical understanding in

the elementary classroom. In R. M. Gillies (Ed.), *Collaborative learning: Developments in research and practice* (pp. 47–68). New York, NY: Nova Science.

*Wenzel, T. J. (2007). Evaluation tools to guide students' peer-assessment and self-assessment in group activities for the lab and classroom. *Journal of Chemical Education, 84*(1), 182-186. doi:10.1021/ed084p182

Wiliam, D. (2009). *Assessment for learning: Why, what and how?* London, UK: University of London.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37* 3–14. doi:10.1016/j.stueduc.2011.03.001

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). New York, NY: Routledge.

*Wu, C., Chanda, E., & Willison, J. (2013). Implementation and outcomes of online self and peer assessment on group based honours research projects. *Assessment & Evaluation in Higher Education, 39*, 21–37. doi:10.1080/02602938.2013.779634

*Yurdabakan, I. (2011). The investigation of peer assessment in primary school cooperative learning groups with respect to gender. *Education 3–13, 39*, 153–169. doi:10.1080/03004270903313608

*Zhang, B., & Ohland, M. W. (2009). How to assign individualized scores on a group project: An empirical evaluation. *Applied Measurement in Education, 22*, 290–308. doi:10.1080/08957340902984075

| Search for articles and dissertations
PsychINFO – 2600 results
ERIC – 2375 results
Libris 135 results | → | Exclusion
4881 search results excluded (were not considered relevant) |

| Titles, keywords and abstracts read
229 articles and dissertations considered relevant | → | Exclusion
107 articles and dissertations excluded (were not within an educational context) |

| Selection
122 articles and dissertations selected that met the inclusion criteria |

| Coding
122 abstracts of articles and dissertations coded | → | Exclusion
90 articles and dissertations excluded (were not within a school context) |

| Additional search for key references
21 articles added
In sum, 143 articles selected | → | Exclusion
53 articles and dissertations excluded during process of thematisation (lacked relevance, were not in the area of group work assessment) |

| Thematisation
90 abstracts thematised | → | Exclusion
7 publications excluded, 5 of them were not in the context of group work and 2 of the dissertations could not be found in full text |

| Reviewing
83 articles and dissertation reviewed in full text |

Figure 1. Search and exclusion process inspired by PRISMA.

Group Work Assessment

The Purpose of Group Work Assessment

Improving group work
– Achieving accountability
– Promoting group work

Promoting learning
– The individual level
– The group level

Giving grades

Emulating real life

What is Assessed in Group Work

Contribution
– Group interaction
– Intellectual
– Workload

Knowledge

Product

Methods of Group Work Assessment

Individual assessment
– Self-assessment
– Portfolio
– Teacher observation
– Combinations of peer assessment

Group assessment
– Group outcome
– Computer-based group assessment

Combined individual and group assessment

Effects and Consequences of Group work Assessment

Effects on the individual
– Learning
– Grades
– Psycho-social effects

Effects on the group
– Cooperation
– Contribution
– Relations

Consequences for the teachers' assessment practice
– Grading
– Obtaining information

Quality of Group Work Assessment

Quality in peer assessment
– Validity and reliability
– Fairness

Quality in methods for individual assessment
– Validity and reliability
– Fairness

Quality in group grading

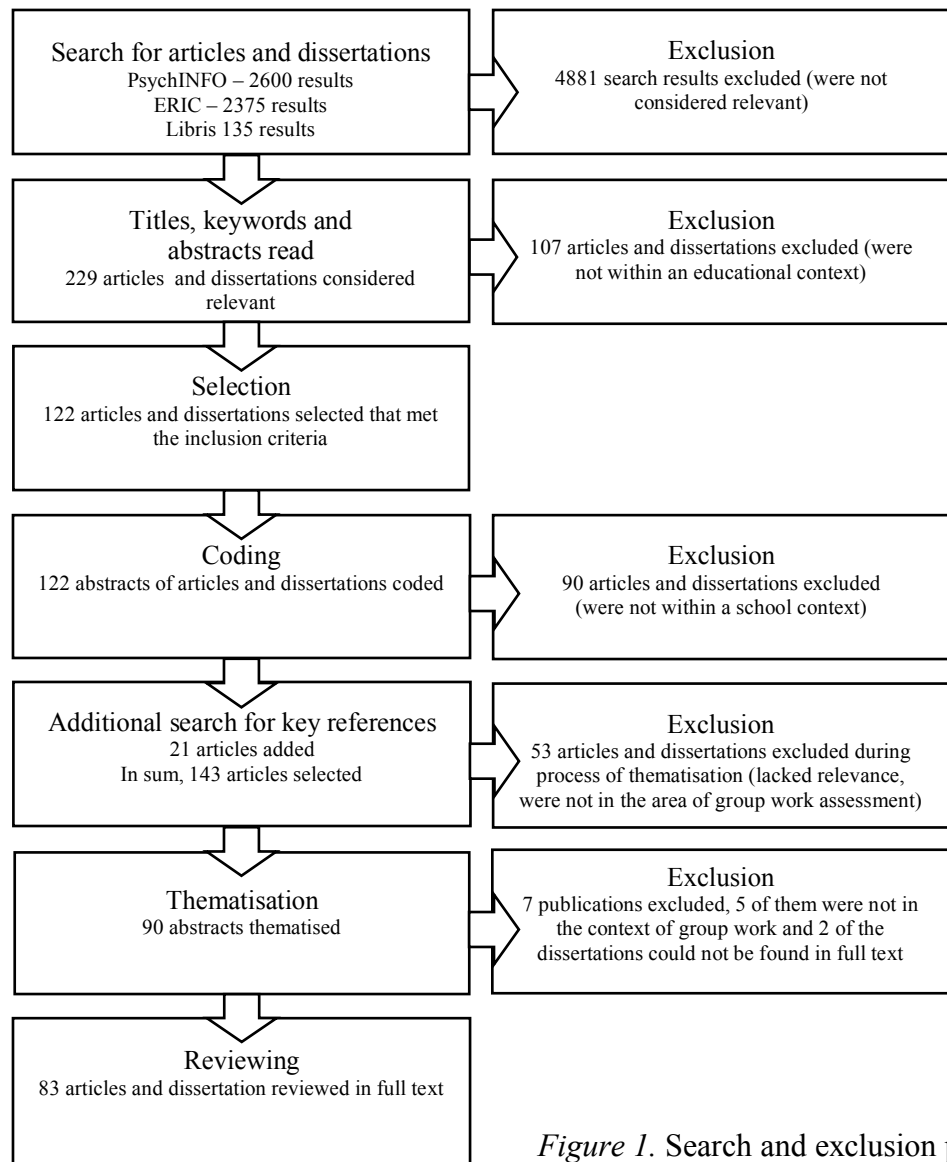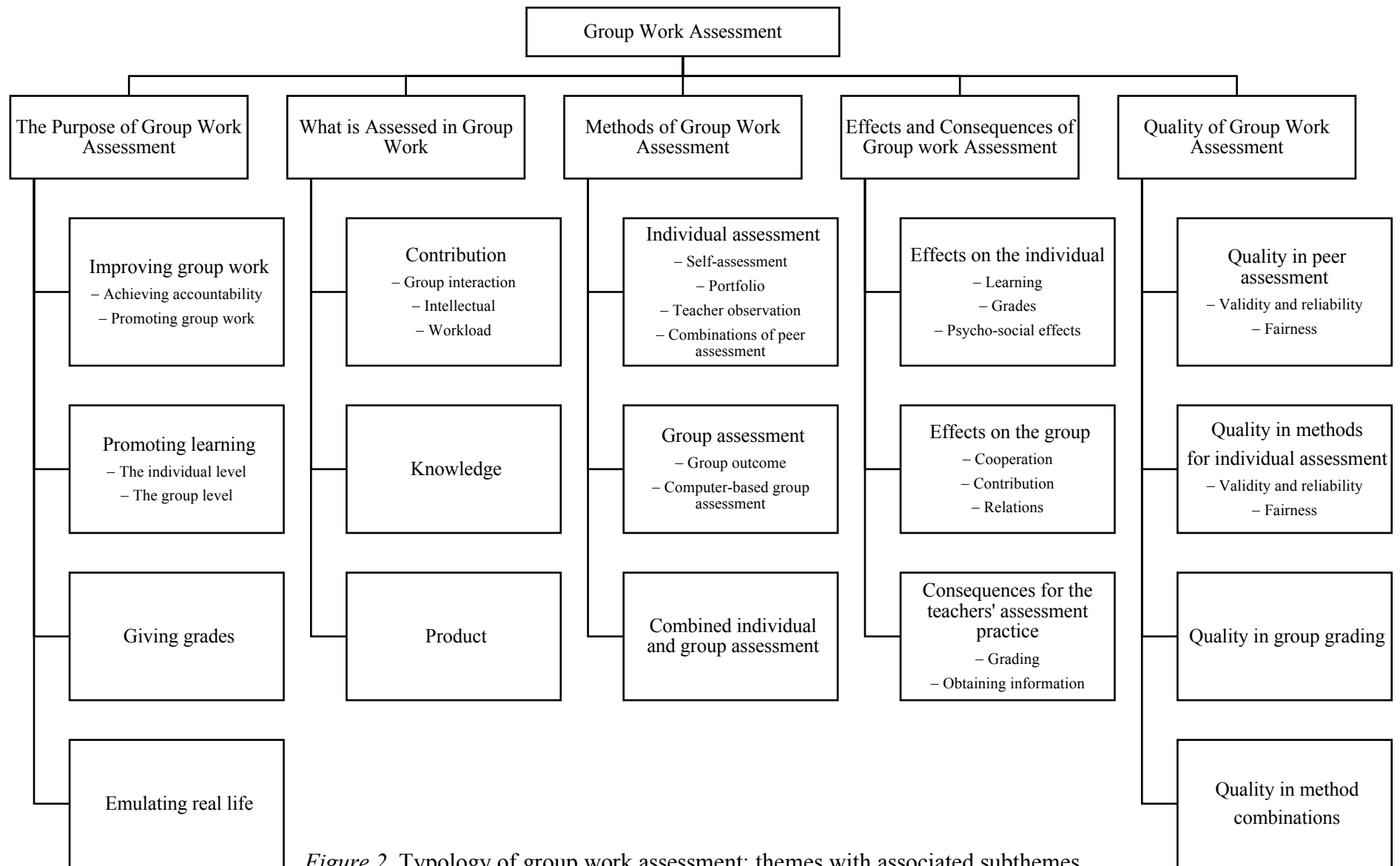Quality in method combinations

Figure 2. Typology of group work assessment; themes with associated subthemes.