

The effects of gender on group work process and achievement: an analysis through self- and peer-assessment

Sachiko Takeda* and Fabian Homberg

Bournemouth University, UK

The importance of teamwork skills as part of employability has been widely acknowledged and accompanied by active research on successful cooperative learning. However, relatively few studies have focused on the effects of gender on students' group work, and only a limited number of empirical studies exist that examine students' group work process and performance through the results of self- and peer-assessment. This study examines the effects of gender on group work process and performance using the self- and peer-assessment results of 1001 students in British higher education formed into 192 groups. The analysis aggregates all measures on the group level in order to examine the overall group performance. Further, a simple regression model is used to capture the effects of group gender compositions. Results suggest that students in gender balanced groups display enhanced collaboration in group work processes. The enhanced collaboration could be associated with less social loafing behaviours and more equitable contributions to the group work. However, the results imply that this cooperative learning environment does not lead to higher student performance. Students' comments allow us to explore possible reasons for this finding. The results also indicate underperformance by all-male groups and reduced collaborative behaviours by solo males in male gender exception groups (i.e., groups consisting of one male student and other members being female). The results thus have implications for the composition of groups. The pedagogical implications of these findings are discussed.

Introduction

It is widely accepted that developing subject-specific technical competence is no longer sufficient for enhancing students' employability through higher education (Cotton, 2001). The use of teamwork in the business environment has steadily increased since the early 1980s (Freeman, 1996) and in most organizations, employees spend significant time working in one or more teams (Katzenbach, 1997; Chapman *et al.*, 2006). Accordingly, employers seek employees who can work effectively within a team environment (Tarricone & Luca, 2002) and expect graduates to have already acquired teamwork skills during their education (Bonanno *et al.*, 1998; Cotton, 2001). This phenomenon highlights the need for higher education institutions to develop students' teamwork skills (Graen *et al.*, 2006) and has led to an increase in the use of group-based learning and assessment in higher education curricula (Chapman *et al.*, 2006). Students across disciplines, but especially those of

*Corresponding author. The Business School, Bournemouth University, EB401 Executive Business Centre, 89 Holdenhurst Road, Bournemouth, BH8 8EB, UK. Email: stakeda@bournemouth.ac.uk

business studies, are often assigned to project teams in their classes (Bolton, 1999; Pitt, 2000).

Accompanying the increased use of group work as a pedagogical tool in curricula are open debates on 'the challenge [of] how to make the group experience as relevant and rewarding as possible for students' (Mello, 1993, p. 253). For example, Johnson and Johnson (1999) list five elements that make cooperation possible: (i) positive interdependence; (ii) individual accountability; (iii) face-to-face promotive interaction; (iv) appropriate use of social skills; and (v) group processing. However, cooperative learning often fails, with social loafing cited as one of the main causes of failure. Social loafing occurs when group members fail to contribute to teamwork, benefit from the efforts of others, and acquire the same rewards as other members of a group. McCorkle *et al.* (1999) report that more than 65% of students identify social loafing as a problem they experienced in group work. Other researchers also suggest that social loafing is a significant factor (Fellenz, 2006; Gammie & Matson, 2007) that affects students' attitudes towards group work (Pfaff & Huddleston, 2003; Stark *et al.*, 2007).

As one effective method of reducing social loafing, many instructors recommend peer-assessment (Cheng & Warren, 2000; Baker, 2008) and empirical studies have evaluated its benefit (Chapman & van Auken, 2001; Pfaff & Huddleston, 2003). Brooks and Ammons (2003) provide evidence for reduced social loafing in group projects when such work is accompanied by peer-assessment. They used peer-evaluations as both summative and formative assessment. Summative assessment is 'concerned with summing up or summarizing the achievement status of a student' (Sadler, 1989, p. 120). It is used to grade students after the course has concluded (Taras, 2007). Formative assessment makes use of judgments about the quality of students' work in order to improve their competences (Sadler, 1989). Using peer-assessment for formative purposes has been recommended (Fellenz, 2006), as peer feedbacks during the process of group work can be expected to generate a student's progress in learning (Druskat & Wolff, 1999; Hernandez, 2002; Baker, 2008). Peer-assessment is especially beneficial in reducing social loafing when it is used as part of the summative assessment. Feichtner and Davis (1984) report that three students in five have the best group experience when peer-evaluation is included in the grading system, as compared to one in three when peer-assessment is not used.

Besides the effects of reducing social loafing in group work processes, the benefits of self- and peer-assessment on students' learning have been pointed out by a number of researchers (e.g., Boud *et al.*, 1999, 2001; Nicol & Macfarlane-Dick, 2006; Nicol, 2010a; Sadler, 2010). Using peer-assessment for summative purposes within a working group bears the possibility of inhibiting cooperation (Boud *et al.*, 1999), which may in some cases lead to a reluctance by academics and students to include it in the grading process (Liu & Carless, 2006; Nicol & Macfarlane-Dick, 2006). However, peer learning accompanied by assessment of its process by students themselves (reflection and self-assessment) helps them develop the vital skills they need in their future professional development and lifelong learning (Boud *et al.*, 1999; Boud, 2001; Nicol & Macfarlane-Dick, 2006). Nicol and Macfarlane-Dick (2006) endorse self-assessment as an effective way of developing learners' capacity for self-reflection. Nicol (2010b) further claims that the ability to evaluate one's peers is an attribute

employers seek in new recruits, and that the development of learners' self-regulation skills requires the regular use of self- and peer-assessment.

Involvement of students in the process of feedback formation is advocated as another crucial aspect of self- and peer-assessment. Sadler (2010) proposes the intensive use of purposeful peer-assessment as a pedagogical strategy for the creation of formatively effective teacher–learner interactions. This symbolizes moving away from the model of 'feedback-as-telling' (Sadler, 2010, p. 548), which solely depends on teachers for the feedback construction. Nicol (2010a) supports this by further emphasizing the significance of peer critiquing as a means to engage students in the construction of feedback.

Despite the increased recognition and use of self- and peer-assessment, and the importance of placing emphasis on learning processes rather than focusing solely on the learning outcomes (Boud *et al.*, 1999), a limited number of empirical studies exist that examine students' group work process—focusing on their collaborative behaviours—through the results of self- and peer-assessment. Further, few studies examine the relation between students' group work process, investigated through self- and peer-assessment results, and group achievements as determined by the final performance marks. The majority of group work behaviour studies are conducted with surveys or questionnaires, and the few empirical studies examining students' self- and peer-assessment of group work focus on evaluating peer-evaluation instruments (Johnson & Smith, 1997; Baker, 2008), reliability of self-ratings (Lejk & Wyvill, 2001; Johnston & Miles, 2004) and peer-ratings (Burke, 1969; Magin, 2001; Ryan *et al.*, 2007), rather than examining students' group work process through the self- and peer-assessment results.

The current study fills this gap and contributes to the research on students' group work by examining the effects of gender on students' group work in regard to both process and achievement, analyzed through self- and peer-assessment which was used as part of a summative assessment. The necessity of more research on the processes underlying the effects of diversity is pointed out by Milliken and Martins (1996), who refer to an organization's balancing act between the affective costs and the cognitive and symbolic benefits of diversity. Student diversity can be related to a wide range of characteristics, such as gender, ethnicity, age, social origin, sexuality or cognitive abilities in general, which might affect a student's attitudes and behaviours in the group work processes. Such characteristics may also affect a student's abilities to contribute to the group work. In particular, perspectives on group work processes—on what constitutes quality contributions to the group work—may vary and thus impact the self- and peer-assessment. This study focuses on only one of the aforementioned categories, gender. Thus, the results need to be interpreted keeping in mind that diversity stemming from other categories might have an additional influence.

Moreover, it needs to be acknowledged that gender can be deconstructed revealing many embedded categories (Lorber, 1993, 1996). Nevertheless, the current study treats gender as a dichotomous variable, as many studies of gender composition in work groups have done (Williams & Meán, 2004), for mainly two reasons. First, the sample students have been identified within the organizational record as male or female. Second, an in-depth investigation of the gender construct is beyond the scope

and purpose of this study, as such an endeavour requires a completely different research design and method.

The current study focuses on the degree of equitable contributions from team members towards the group work—achieved either through perceived absence of social loafing behaviours or through an absence of ‘outstanding members’ in the group, i.e., someone who is perceived by all his/her group peers to have made greater contributions than other members. The degree of equitable contributions from group members is judged by variations in the ratings students gave each other within a group in their self- and peer-assessments. Students were instructed to give a rating of 6 (out of 10) to those group members whose contribution to the group work was considered roughly equivalent to that of everyone else. They were instructed to give ratings of more than six to those members whose contributions to the group work were perceived to be greater than those of most members of the group, and give ratings of less than six to members whose contributions were considered to be smaller.

When members perceive no social loafing behaviours in the group and no members are perceived as having contributed more than others, the variation of self- and peer-assessment ratings becomes zero, with every member of the group giving each other a rating of six. This indicates that group members perceive that equitable contributions have been made towards the group work. On the other hand, when the group work process is either affected by social loafing behaviours or led by larger contributions by particular members, the variation of self- and peer-assessment of the group increases, indicating the existence of perception in the group that less equitable contributions have been made by the group members towards the group work process. Qualitative comments by students which accompanied the self- and peer-assessment ratings in two additional modules are used to support the quantitative analyses.

Group diversity and its effects on group work

The effect of team member diversity on team performance is an actively studied area (Maznevski, 1994; Milliken & Martins, 1996) both in workplace settings and educational institution settings. Research has been conducted investigating the influence of diversity in teams, such as top management team composition (Carpenter *et al.*, 2004; Nielsen, 2010), boardroom composition (Adams & Ferreira, 2009; Miller & Triana, 2009) or work teams composition (Joshi *et al.*, 2011), including those focusing on ethnic diversity (Earley, 1989) and gender diversity (Wood, 1987; Carli, 2001). At educational institution settings, the effect of ethnic diversity in student teams is intensely researched (e.g., Watson *et al.*, 1993; McLeod *et al.*, 1996; Paulus *et al.*, 2005). A few comprehensive studies also exist focusing exclusively on the effect of gender diversity in students’ team performance (Sormunen-Jones *et al.*, 2000; Orlitzky & Benjamin, 2003).

A review by Milliken and Martins (1996) reveals the literatures’ broad overall agreement that diversity in observable attributes such as race/ethnic background, nationality, gender and age prevents smooth group integration in the team work process. Maznevski (1994) introduces the general conclusion of previous research on group performance that diverse groups perform less well than homogeneous ones do, although this disadvantage of diversity can be moderated or turned into advantage

with better communication. However, the evidence researchers have provided is not consistent. Jehn *et al.* (1999) found diversity based on age and gender positively affects worker morale, including satisfaction, commitment and perceived performance. The findings of Robbins and Fredendall (2001), on the other hand, suggest that homogeneity is positively related to team success and motivation.

Findings are more consistent regarding studies that focus on gender diversity in teams. Orlitzky and Benjamin's (2003) survey of 138 students shows that mixed-gender groups outperform more homogeneous groups. Wood's (1987) meta-analytic review suggests a tendency for mixed-sex groups to outperform same-sex ones, though the tendency cannot be determined as significant. A meta-analytic review by Carli (2001) reveals that in gender diverse groups, male members exert stronger influence than female members and contributions by male members receive more attention from other group members than contributions by female members. This tendency strengthens in gender imbalanced groups, especially in groups with only one male or one female, putting women in minority at a disadvantage. Having a gender balance in groups, on the other hand, enhances women's influence and weakens the inequality of influence between the genders. This consequently leads to a change of the nature of the group members' interaction, creating mutual support and agreeableness within the groups.

Thus research outcomes of the effect of diversity on group process and performance widely vary. When developing our hypotheses, we focus on the results of meta-analytic reviews. Considering the review by Maznevski (1994), suggesting that diverse groups tend to be outperformed by homogenous ones, and bearing in mind Wood's (1987) meta-analytic results pointing towards insignificant results, we derive the first hypothesis:

- Hypothesis 1: Gender diversity in group component leads to lower final performance marks.

Furthermore, we consider the effects of gender diversity on group work process. The review by Milliken and Martins (1996) suggests negative effects of diversity on the group work process. In contrast, in a meta-analysis Carli (2001) found that male team member's dominance is mitigated by having a gender balance in groups, which leads to the creation of mutual support and equality of influence between genders within the groups. Carli's (2001) analysis focuses on gender diversity, while Milliken and Martins (1996) consider diversity in various observable attributes. As this research focuses on the effects of gender diversity, the second hypothesis is formed as follows:

- Hypothesis 2: Members in gender balanced groups make more equitable contributions in the group work process.

Gender traits and their effects on group work

When examining the effect of gender composition on the group work process and achievement, two perspectives regarding gender traits need to be considered. One is the tendency of female students to academically outperform male students in overall school curriculum achievements (Bernard, 1997; Want *et al.*, 2002; Tinklin, 2003) as

well as achievements in some specific areas, such as reading (Topping *et al.*, 2008). Female learners' higher achievement in school coursework may also be contributing to their success in national examinations, such as GCSEs in England, Wales and Northern Ireland (Elwood, 2005). Detailed analyses reveal more complex subject-specific gender success rates in these examinations (Elwood & Murphy, 2002), as well as gender differences in attitudinal and academic developments through the education system (Murphy & Elwood, 1988).

Female learners' outperformance over male learners is often attributed to the different attitudes towards school work displayed by the genders from early stages of education; whereas boys experience pressure to appear 'cool' and not care about school work, girls experience more accepting attitudes from peers for them to work hard (Warrington *et al.*, 2000). Under such conditions, females take school more seriously, are better prepared, better organized and exhibit more respectful and cooperative attitudes, while boys are ill-prepared, competitive, disruptive and less attentive (Tinklin, 2003).

The study by Warin and Dempster (2007) reports that in line with the necessity of appearing 'cool' at schools, male students at university try to give an impression of being 'laddish', associated with such features as heavy drinking and loudness. This is part of their temporary efforts to fit in and to be accepted amongst peer students and limited to the early days until they find more authentic friendships. According to the study, however, appearing less committed to academic practices is not included in these leanings towards apparent 'laddishness'. This is because by entering university, male students accept an element of academic effort in themselves and in their male peer group.

Nevertheless, even without the peer pressure to appear less serious about academic work, male students continue to display the tendencies of less commitment and lower achievement in higher education. Woodfield *et al.* (2006) found that male undergraduate students, with significantly higher absence rates, were outperformed by female students in final degree outcomes. This leads to the third hypothesis of the current study:

- Hypothesis 3: All-male groups achieve lower final performance marks than groups of any other composition.

The other perspective is of a more general gender stereotypical trait. A number of studies suggest that women's traits include agreeableness, extraversion, solidarity display, understanding, helpfulness, selflessness and nurturance, while males tend to be self-assertive, controlling, aggressive and dominant (Eagly, 1987; Feingold, 1994; Budaev, 1999; Costa *et al.*, 2001). Although these are stereotypic traits that need to be considered with caution, female traits are, in general, more in accordance with the aforementioned elements listed by Johnson and Johnson (1999). These elements make cooperation possible, especially with respect to positive interdependence, face-to-face promotive interaction and appropriate use of social skills. The commitment female students show toward their schoolwork (Warrington *et al.*, 2000; Tinklin, 2003; Woodfield *et al.*, 2006) is also in line with the elements of individual accountability and contributes to the fulfilment of the remaining element of group processing that further makes cooperation possible (Johnson & Johnson, 1999).

Moreover, Johnson and Smith (1997) found female students were rated higher than males on traits such as effort, cooperation and initiative that, according to Johnson and Smith (1997), are desirable behaviour traits for group cohesiveness leading to success. Warrington *et al.*'s (2000) study suggests that female students have higher communication skills. Male students, in comparison, are less inclined to use cooperative talk/discussion and unwilling to collaborate to learn, being less of team players. Bernard (1997) also showed that boys found it hard to cooperate in groups and were poor team players. These studies suggest female students tend to have a better experience of the group work process with peer female students than with male students. The few studies that examine the results of peer-assessment from gender perspectives also report female students scoring higher than male students on peer-assessment (Johnson & Smith, 1997; Baker, 2008).

Furthermore, Bernard's (1997) study suggests an all-male environment tends to compound the least attractive aspects of male attitudes and behaviour, including lack of cooperative elements in male traits. This leads to our fourth hypothesis:

- Hypothesis 4: All-male groups experience a more troublesome group work process with less equitable contributions from members.

Gender exception groups and group work

Sormunen-Jones *et al.* (2000) define 'gender exception groups' as groups in which all members are one gender except one of the opposite gender. They found that the 'gender exception groups' achieved lower scores in group writing projects in content, organization, style, and mechanics, as well as in the total achievement score, compared to same gender or mixed gender groups. Studying 486 work groups across varying jobs and organizations, Sackett *et al.* (1991) found that when women composed less than 20% of the group, the women's performance was rated lower than men's. When women make up more than 50% of the group, their performance is rated higher than men's. They did not find the same pattern when men composed less than 20% of the group.

Gammie and Matson (2007) find significant effects, originating from gender differences, in efforts taken to compensate for underperformance in the group; more female students than male students report having undertaken additional effort. This may suggest a disadvantage for female students when they form the minority in a group. The fact that they are unable to share the additional work taken up to compensate the underperformance of certain members of the group with other female students might create such a disadvantage. Kaenzig *et al.* (2007) found that female students in business studies reported more negative experiences with group work assignments than did male students. This may be related to their having to undertake additional efforts in the group work.

Contrary to the disadvantageous position women face in minority, the study by Craig and Sherif (1986) found that compared to groups with two men/two women or with three men/one women, men exerted a larger amount of influence over other members and groups' decisions when they were in a minority of one in a group. Further, the aforementioned Carli's (2001) meta-analysis supports these findings. Thus,

previous studies suggest that being the only one gender in a group creates a more influential position for male gender exception members, yet a disadvantageous position for female ones. This indicates that the members in gender exception groups tend to evaluate the contributions by the solo male or solo female members to be less equitable. This, combined with Sormunen-Jones *et al.*'s (2000) findings, leads to our fifth and sixth hypotheses:

- Hypothesis 5: Gender exception groups achieve lower final performance marks.
- Hypothesis 6: Members of gender exception groups evaluate the members' contributions to the group work as less equitable.

Sample

The data used in the current study were collected over five years from the academic years of 2005–2006 to 2009–2010 from first year undergraduate business students following an international management module in a higher education institution located in the UK. The module ran for 22 weeks each year, but the group work of the study was used to assess the students' learning for the first half of the module only. Besides the group work, a multiple-choice test was used to assess students' cognitive understanding of the subject. The results of the group work accounted for 25% of the mark for the entire module. Across five years, 1001 students (601 male and 400 female) were allocated by a tutor into 192 groups (4 with 3 members, 22 groups with 4 members, 100 groups with 5 members, 61 groups with 6 members, 5 groups with 7 members). These groups were assigned the task of presenting information on cultural awareness when doing business with people of a country of the groups' choice.

After the presentations, each student submitted self- and peer-assessments in which he/she evaluated on a scale from 1 to 10 the contributions made by himself/herself and each of his/her respective group members. Although students were free to give any rating from 1 to 10 to any member, giving a rating higher than eight or lower than four required prior consultation with the tutors. The ratings submitted were not revealed to other members of the groups. For using the self- and peer-assessment for summative purposes—i.e., in order to reflect the ratings each student received from their peers (and themselves) for their contributions to the group work on their final assessment marks—the presentation mark (hereafter: performance mark) given to the group by the tutor was then modified using the self- and peer-assessment ratings with the following calculation. First, the average of the ratings each student received from his/her peer group members and self-assessment rating was divided by the average of all self- and peer-assessment ratings given within the group. The result was then multiplied by the group performance mark, representing the student's individual final mark for the group presentation assignment.

As mentioned above, the sample on which the subsequent analyses are based contains 192 groups of students. However, a drawback of the dataset is that in this particular module students were not required to provide qualitative comments justifying the marks they attributed to themselves and their peers. Nonetheless, to achieve a better understanding of students' rating behaviours and group work processes, such

qualitative comments are very useful. Therefore, a second set of data originating from two different modules on the same course is examined.

In these two modules students were asked to provide, in addition to their self- and peer-ratings, justifications for each rating and explain each peer's, as well as his/her own, contribution to the group work in 100 to 150 words. With each student providing comments on themselves and every other group member, 2081 comments were collected from 319 students forming 66 groups over the two academic years of 2010–2011 and 2011–2012.

While the instructions for the self- and peer-assessment ratings given in these two modules remained identical to those in the sample used for the subsequent quantitative analyses, the instructions for the group work, i.e., the tasks driving the final performance marks, were different. The task in one of the two additional modules consisted of two parts: the first part was identical to the task of the module providing the main data for the quantitative analysis of this study. The second part asked students to present a cross-cultural training design for an expatriate who would move abroad in the near future. The task of the other additional module was, again, to present information on cross-cultural awareness when doing business, though a stronger focus on the applicability of a management theory in two different countries characterized the task. All three modules used the same cross-cultural theories as the basis, and the same lecturers were involved in the marking of the presentation assignments. Yet, the tasks were not identical, and it was therefore inappropriate to pool the datasets from these different modules, mainly for the incompatibility of the final performance marks. Nevertheless, following the same instructions, the self- and peer-assessment rating results of these additional modules also contribute to this study's examination of students' collaborative behaviours, and the accompanying comments help us better understand the students' group work processes.

Methods

In order to examine the effect of the gender component in the group based on the self- and peer-assessment rating variations and group presentation marks, we categorize all groups into six different categories. All 192 groups are categorized: (1) all male; (2) all female; (3) all male except for one female; (4) all female except for one male; (5) balanced gender (groups with an equal number of male/female); and (6) all other gender component. See Table 1 for a breakdown of the number of each type of group.

Table 1. Summary of group types and group sizes

Group type	Gender composition	N
1	All male (M)	23
2	All female (F)	1
3	All male + 1 female (M + 1F)	22
4	All female + 1 male (F + 1M)	10
5	Balanced	40
6	Others	96

Variables

For the first part of the analysis it is necessary to aggregate all measures on the group level in order to examine the overall group performance. We then analyze gender effects using two commonly used indices in management research, gender quotas and the Blau Index. The second part of the analysis captures the group work contribution behaviours of individual students. Thus this part of the analysis relies on individual student data, not on group level data.

Dependent variable. In our model *performance* indicates the dependent variable. We use two different performance measures. First, *performance mark* measures overall group achievement according to the mark attributed to each group's presentation. Second, we employ the *self- and peer-assessment variation* to capture the group work process. We use the standard deviation of the self- and peer-assessment marks to capture the aggregated group level effect of the marks attributed to each individual group member. As mentioned before, a group work process with equitable contributions from its members is reflected by zero variation.

Independent variables. In order to assess a direct gender effect we calculate the *gender quota* as a percentage of female group members. In order to capture more general effects relating to *diversity* we use the Blau Index as a measure for group heterogeneity. The Blau Index is a widely used diversity index in management research (Carpenter, 2002; Nielsen, 2009; Finkelstein et al., 2009) and is calculated according to Equation 1:

$$(1) \quad B = [1 - \sum(P_i)^2],$$

with p being equal to the percentage of the group members in the i -th category (i.e., gender).

Analysis

Group performance. We calculate effect sizes and the respective confidence interval for each group category compared to the remainder of the sample. Effect sizes are standard measures in meta-analyses (Lipsey & Wilson, 2001; Stanley, 2001; Ellis, 2010) and benefit from the fact that they are comparable across studies. The latter is due to the fact that they are calculated on a dimensionless scale. Conceptually, this approach of analysis can be compared to an experimental one. Each one of our group categories can be thought of as being the 'treatment' group, whereas the groups constituting the rest of the sample are the untreated ones. We ultimately analyze differences between these categories.

The effect size we selected is Hedges' g , which is recommended where groups differ in size (Ellis, 2010, p. 11). This is clearly the case in our sample. We calculate Hedges' g for each group category and compare it to the remainder of the full sample. Hedges' g (Hedges, 1981; Ellis, 2010) is calculated according to Equation 2:

$$(2) \text{ Hedges' } g = \frac{M_1 - M_2}{SD_{pooled}^*}$$

In equation 2 M_1 and M_2 indicate the means of Group 1 and Group 2 respectively. SD_{pooled}^* indicates the pooled and weighted standard deviation of the two groups. Weightings occur according to group size. Hedges' g must be interpreted in terms of standard deviation units, i.e., higher scores indicate larger effects.

In order to ease interpretation of the resulting values, we convert Hedges' g to the probability of superiority (Grissom, 1994; see also McGraw & Wong, 1992 who refer to it as the 'common language effect size'). The probability of superiority is helpful because it transforms the effect size (i.e., Hedges' g in this case) reflecting the difference between two groups into a probability that is easily interpreted. The 'probability of superiority' indicates the probability that a randomly selected value from Group 1 will be greater than a randomly selected value from Group 2 (Grissom, 1994; Ellis, 2010). According to Grissom (1994) probabilities of 0.56 ($g = 0.20$) are classified as small, probabilities of 0.64 ($g = 0.50$) as medium and probabilities of 0.71 ($g = 0.80$) are considered large. These values reflect the standards for effect sizes developed by Cohen (1988).

Finally, we use a standard regression model to analyze effects of diversity on group performance. We estimate a standard OLS model with robust standard errors using group size and the absolute number of female group members as control variables. Equation 3 presents the estimated model:

$$(3) \text{ Performance} = \alpha + \beta_1 \text{Diversity} + \beta_2 \text{Gender} + \beta_3 \# \text{Female} + \beta_4 \text{size} + \varepsilon_i$$

Table 2 below presents the correlations of all variables.

Qualitative analysis of comments. We added the dataset of qualitative comments to explore potential, hardly observable drivers of group work processes. In preparation for the analysis, each group in the dataset was given an individual alpha-numeric identifier, and each member within a group was given a numeric identifier. These

Table 2. Correlations

	#Females	#Males	#Members	%Female	%Male	Diversity	Performance mark	SPA variation
#Females								
#Males	-0.77							
#Members	0.24	0.42						
%Female	0.95	-0.89	-0.00					
%Male	-0.95	0.89	0.00	-1.00				
Diversity	0.71	-0.56	0.15	0.70	-0.70			
Performance mark	-0.01	0.04	0.04	-0.00	0.00	0.00		
SPA variation	-0.08	0.01	-0.10	-0.05	0.05	-0.14	-0.13	

Notes: All correlations > .04 are significant at the 5% level.

identifiers are unique to this study, securing the anonymity of the students. In analyzing comments, we followed an inductive approach letting categories emerge from the data (Waring, 2010), which requires some judgment on behalf of the researchers (Jenkins & Johnson, 1997). An inductive approach to the qualitative data analysis can be used to ‘establish clear links between the evaluation or research objectives and the summary findings derived from the raw data’ (Thomas, 2006, p. 237). A first screening of the qualitative data revealed the salient issue of outstanding members in the groups, which later became our overarching category and is also reflected in the quantitative analyses. A second, more in depth reading of the data enabled us to further refine this theme in order to understand what constitutes such members. As a result, the categories ‘leader’, ‘creative ideas’ and ‘helping and doing the work of others’ were created and taken as a basis for the subsequent analysis, which is discussed in the final sections.

Results

Self- and peer-assessment variation

As a first step in the analysis, the effect of the gender component in the group on the degree of equitable contributions from group members is evaluated using the self- and peer-assessment rating variations. See Table 1 for a breakdown of the number of each type of group. The variation of the ratings within the different types of groups is shown in Table 3. A visual inspection of box-plots indicates that the gender balanced groups have notably lower variation compared to other types of groups.¹

In order to capture the group work process, we calculate Hedges’ *g* for the standard deviation of the self- and peer-assessment marks. Table 3 displays the results. In particular the results show that gender balanced groups have lower variation in self- and peer-assessment scores confirming the results of the box-plot inspection. The probability of superiority takes medium values for group categories 4 and 5, hinting towards a positive effect stemming from group diversity.

Performance mark

Of utmost relevance for students is the final performance mark attributed to each group. The effect sizes of the different group categories reveal that gender balanced

Table 3. Hedges’ *g* for self- and peer-assessment (SPA)

Group category	N	Hedges’ <i>g</i>	[95% Conf.Interval]	SE	Prob. of Sup.	
1 – all male	23	0.401	–0.037	0.838	0.223	0.6114
3 – all male +1f	22	0.005	–0.439	0.449	0.227	0.5000
4 – all female +1m	10	0.67	0.03	1.311	0.327	0.6822
5 – balanced	40	–0.518	–0.87	–0.165	0.179	0.6434
6 – others	96	0.051	–0.232	0.334	0.144	0.5141
2 – all female	1	excluded				
Total	192					

Table 4. Hedges' g for performance mark

Group category	N	Hedges' g	[95%	Conf.Interval]	SE	Prob. of Sup.
1 – all male	23	–0.17	–0.606	0.266	0.222	0.5478
3 – all male +1f	22	0.376	–0.069	0.822	0.227	0.6059
4 – all female +1m	10	0.202	–0.435	0.839	0.325	0.5562
5 – balanced	40	–0.002	–0.35	0.346	0.178	0.5000
6 – others	96	0.111	–0.172	0.394	0.144	0.5310
2 – all female	1					
Total	192					

Table 5. Regression results

VARIABLES	Dependent variable: performance			
	Model 1 (SPA variation)		Model 2 (performance mark)	
	β	t	β	t
Diversity	–0.976** (–0.488)	–2.00	–15.60* (–8.418)	–1.85
% Female	0.30 (–1.65)	0.18	–15.2 (–24.41)	–0.62
# Female	0.0027 (–0.314)	0.00	2.74 (–4.655)	0.59
Group size	–0.0432 (–0.127)	–0.34	–1.566 (–2.028)	–0.77
Observations	168		168	
R-squared	0.039		0.021	

Notes: Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

groups display a slightly lower performance mark than groups comprising minorities (Table 4). It seems that low diversity groups, i.e., categories 3 and 4, performed best. The performance mark is lower when the group consists of male students only. However, all effects remain small.

Regression models

Table 5 displays the results of our regression analysis. We estimate two models using the two different performance measures—self- and peer-assessment rating variation and performance mark—as dependent variables. Since we are ultimately interested in the effects of group diversity, we include only those observations in the regression models which have a diversity index value larger than zero. The results indicate that the absolute number of female students in each group, the overall group size and the gender quota do not exhibit significant effects on our performance measures. On the

other hand, the diversity variable exhibits some notable effects. Both in Models 1 and 2 the coefficient on the diversity variable is significant and negative, i.e., the higher the diversity in a group the lower is the rating variation in the self- and peer-assessment exercise (Model 1) and the lower is the performance mark (Model 2).

Discussion

Neither the results based on the effect size, nor those based on the regression analyses, provide strong evidence for gender effects on the final outcome of group work. The finding that gender balanced groups have a tendency to display lower performance marks is supported by the negative significant coefficient on the performance marks, hence supporting Hypothesis 1. These results compete with the findings of Orlitzky and Benjamin (2003), who find stronger positive effects of gender heterogeneity in group member component on students' group performance.

The most striking finding of this study relates to work process which seems to be strongly affected by gender composition of the various groups. The results demonstrate that members in gender balanced groups make more equitable contributions to the group work compared to those in gender imbalanced groups, supporting Hypothesis 2. This is reflected in the results of the self- and peer-assessment rating variations which assume medium effect size values in these cases (see Table 3), rather than small ones. Further, the significant diversity coefficient in Model 1 provides similar support. Robbins and Fredendall (2001) suggest a connection between the equity of team member contribution and higher standards established by the team. Fall *et al.* (2000) show that even small amounts of peer collaboration can significantly help students' learning and performance. The analysis by Baines *et al.* (2007) indicates that involvement in the group work programme has a positive effect on learners' learning outcomes. These results emphasize the importance of a collaborative team climate in the process of cooperative learning for the achievement of better learning results.

Nevertheless, the findings of the current study indicate that the equity of group member contributions does not necessarily have positive effects on the students' learning outcomes. Despite the more equitable contributions made by members, gender balanced groups achieved lower performance marks. Closer examination of the raw data of self- and peer-assessment results and students' comments implies that one of the key contributing factors to this phenomenon may be the existence or absence of a member with outstanding contributions within a group (hereafter called 'outstanding member').

Out of 192 groups analyzed in the current study, 58 (30.2%) groups had members who (1) received at least 0.5 points higher average of self- and peer-assessment ratings than the group average; and (2) received the rating of seven or higher from all other members, hence being acknowledged to have made larger contributions to the group work than average (for which students were instructed to give a rating of six) by all other members. Of these outstanding members, 29 were male and the other 29 were female (4.8% of all male students and 7.3% of all female students).

Amongst our sample of 192 groups, outstanding members exist most frequently in the groups that achieved the highest category of performance mark (13 out of 21

groups or 61.9%), compared to 26.3% amongst the lower categories altogether. While this indicates the role these outstanding members play in carrying the groups forward to achieve the top performance marks, what is noticeable is their absence in the gender balanced groups. Whereas 30.2% (N = 58) of all groups had outstanding members, they existed in only 17.5% (N = 7) of gender balanced groups (N = 40). This is the lowest amongst all group types and may have contributed at least partially towards their lower variation of self- and peer-assessment and lower performance marks.

As a next step it is interesting to investigate how the nature of 'outstanding members' is reflected in the dataset of students' comments.² Students' comments accompanying self- and peer-assessment in other modules reveal that such outstanding members are considered as group leaders by their peers. The word 'leader' was used in describing the contributions of the majority (89.5%) of outstanding members by at least one of their peers. Included in these comments were:

She led the group and was the key driver to the success of the presentation. (Group B1, Member 5)

He was our leader and led us to the right direction. (Group N3, Member 6)

He took the leading role and contributed a lot to ensure the presentation was the best that it could be... (Group B3, Member 3)

He almost controlled the whole process of assignment and the meetings as a leader. (Group O3, Member 1)

She was a real leader and helped the group cohesion. (Group B1, Member 1)

He was a genuine motivator in this group, perhaps as well as a group leader. (Group P1, Member 4)

Her remarkable leadership and organizational skills helped us to stay organized... (Group P2, Member 2)

The other two comments most frequently made in evaluating the contributions of the outstanding members related to the fact that they provided creative ideas (78.9% of the outstanding members received this comment from at least one of their peers) and that they helped and did the work of others (mentioned regarding 68.4% of the outstanding members). Students wrote:

One of our group members did not do anything and she was the one who covered it. (Group F2, Member 6)

I got help from her all the time. (Group J2, Member 4)

She basically designed the whole part on her own, since the other members seemed disinterested and a little less able... (Group O3, Member 4)

She was an asset to the group always helping others, bringing new ideas... (Group P2, Member 3)

He has contributed with really good ideas... (Group A1, Member 1)

He was very creative in his work, he did offer a lot of great ideas... (Group A3, Member 4)

Hedges' *g* for performance mark (Table 4) indicates the tendency that all-male groups achieve lower performance marks than groups of other compositions, which would be in line with Hypothesis 3. However, the effects remain small; our findings cannot establish support for Hypothesis 3. Similarly, with small effects, our findings suggest that the self- and peer-assessment rating variations (Table 3) for all-male groups are higher than groups of other compositions except for male gender exception groups. We interpret this as providing only limited support for Hypothesis 4.

Nevertheless, closer observation of individual peer-assessment ratings suggest support for the previous studies' findings that boys/male students display a tendency to be less committed to their school/course work (Tinklin, 2003; Woodfield *et al.*, 2006) and less co-operative in groups, being less of team players (Bernard, 1997; Warrington *et al.*, 2000). Amongst the 59 students in the current study sample who received the rating of four (the lowest students are allowed to score without previous consultation with the tutors) or lower from their peers, 47 were male. This represents 8% of all male students in the sample whereas only 3% ($N = 12$) of female students received such low ratings. The comments accompanying the peer-assessment rating of four in the qualitative dataset suggest the most common reasons for students to give such low ratings to their peers are (1) not attending meetings—mentioned in 68% of the comments; (2) doing nothing—mentioned in 51% of the comments; and (3) letting others do their work—mentioned in 48.9% of the comments. One student called the case 'a clear example of the free rider problem, or what some may call "social loafing"' (Group A3, Member 2). Other comments included:

She utterly did nothing... (Group F2, Member 5)

For the whole time he was relying on the others to do his work. (Group F2, Member 3)

He failed to turn up to several meetings... and generally seemed disinterested during the meetings that he did attend... He was certainly happy to sit back and let others do the work. (Group O2, Member 2)

He generally seemed unenthusiastic and to not care about the outcome of the task or the course in general. (Group O3, Member 2)

He was generally unavailable for meetings... He failed to show any interest or initiative at all. (Group O2, Member 3)

Bernard (1997) suggests that an 'all-male environment' compounds the least attractive aspects of male traits, including being a poor team player with less cooperative attitudes in groups and being less committed, leading to their lower academic achievements. However, according to Bernard (1997), mixed-gender environments mitigate these traits. This, considered jointly with the tendency of all-male groups to achieve lower outcomes, suggested by the results of the current study, may imply that the formation of all-male groups is not desirable from the point of view of students' learning through collaborative group work.

The results for performance marks (Table 4) indicate that the performance marks achieved by both gender exception groups are higher than groups of other compositions including gender balanced groups, although the effects are again not significant. Hence Hypothesis 5 is rejected. This contrasts the findings of Sormunen-Jones *et al.* (2000) who found that gender exception groups achieved lower performance compared to the same or balanced gender groups.

Although the tendency cannot be determined as statistically significant, gender exception groups with one male demonstrate a tendency for having higher variation in the self- and peer-assessment ratings than other groups (see Table 3). This finding indicates support for Hypothesis 6. However, the results of the current study display this tendency only with male gender exception groups and not with female gender exception groups.

The high variation in the self- and peer-assessment ratings in male gender exception groups was partly caused by the contributions of the solo males in groups being rated lower by their female group members.³ The same pattern was not found with the sole female students. This result is partially in line with previous findings of Tsui *et al.* (1996). They report the relationship between being different in gender and commitment to be negative for men and positive for women.

Further, this result is contrary to the findings of Sackett *et al.* (1991) who reported that the undervaluation of performance did not occur when men composed less than 20% of the group. At the same time, the lower ratings of solo male members can be interpreted as a result of other female group members giving higher ratings to each other's performance. This would support the findings of Sackett *et al.* (1991), which reported when women formed majority in groups, their performance was rated higher than men's. The low ratings of solo male members also contradict the findings of a meta-analytic review by Carli (2001) which revealed that gender exception groups strengthen the tendency that male members exert stronger influence and contributions by them receive more attention from other group members. According to the results presented here, those who are disadvantaged by being a minority are not female, but male members. Furthermore, in our sample, 6 out of 10 male gender exception groups had outstanding members whose contributions were highly evaluated by their peers, and all six of them were female.

The study by Craig and Sherif (1986) suggests that the low rating of solo male members in male gender exception groups could have been a result of their exerting a disproportionately strong influence over other members and group decisions—which was unwelcomed by their female peers. However, the examination of the comments accompanying the peer-assessment reveals that low ratings are rarely given for reasons

other than lack of commitment. There was one case of a student who attended all meetings, yet received the low rating of four from four of his six group members for being 'aggressive' in discussions (Group M1, Member 5) and 'not listening to others' opinions' (Group M1, Member 6). However, this was a gender diverse group with four male and three female students, and the decisive factor for the low ratings was ultimately his lack of contributions explained by his peers as: 'he did not do anything much' (Group M1, Member 3). The low ratings given to the male students in male gender exception groups can therefore be assumed to be the results of sheer lack of commitment by the solo male members, which was, as suggested by the study of Gammie and Matson (2007) and students' comments, compensated by other female members of the group.

Implications for group work as a pedagogical tool

Results of the current study suggest that the formation of male gender exception groups should be especially avoided, considering the learning experience of the sole male members and its implications for the rest of the group members. The results further suggest that, when considering students' learning process and achievement, the formation of all-male groups also may not be desirable.

This has important implications for the ways groups are allocated. As a method of group allocation when using group work as a pedagogical tool, some researchers and practitioners recommend letting students choose their own group members (self-selection) over random assignment (Bacon *et al.*, 1999; Chapman *et al.*, 2006) or instructor assignment (Ciani *et al.*, 2008). Others advocate the advantage of instructor assignment over self-selection (Feichtner & Davis, 1984; Connerley & Mael, 2001). Among all three methods, self-selection is most commonly used (Decker, 1995). However, as warned by Bacon *et al.* (2001), self-selection bears the danger of creating over-homogeneity in the team.

This point was not examined with the current study data because the sample groups were not self-selected by the students. We, therefore, carried out an additional investigation using the dataset collected from two modules from first and second year undergraduate studies, which are different from those used thus far in the current study. The results showed that when self-selecting group members, 27.4% of male students chose to form all-male groups and 32.6% went into gender exception groups, while only 5.2% formed gender balanced groups with equal numbers of each gender. This indicates that given choices, 60% of male students went into the types of groups the findings of the current study suggest to avoid. Amongst female students, 11.2% formed all female groups, 38.3% went into gender exception groups, and 5.6% formed gender balanced groups. This result supports the above-mentioned warning of Bacon *et al.* (2001). Considering the importance of learning through the experience of group work, the findings of the current study support the advocacy of Feichtner and Davis (1984) that instructors assign students into heterogeneous groups or suggest that some measure be taken to ensure students form gender diverse groups.

Limitations and future research

We have to acknowledge some limitations present in this study which can be used as triggers for future research. First, as pointed out in the beginning, we focus on one single diversity category, keeping in mind that various others are present. The dataset used here displays little variability in terms of age (more than 90% being of the age between 17 and 20) and nationality (95% being UK students) and does not allow us to identify other student characteristics, such as religion, ethnicity, social class, sexuality or being a member of an existing group of friends. These characteristics could have affected the students' attitudes and behaviours in the group work processes or their ability to contribute to the group work as well as their perspectives as to what constituted quality contributions. Thus, future research could try to investigate how student characteristics other than gender affect their perspectives on group work processes as well as their group work participation behaviours.

Second, we employ the gender categories of male and female in this study. However, categorizing gender as a polarized, dichotomous variable in research poses a danger of oversimplification, missing nuanced detail (Young, 1994; Knaak, 2004). Deconstruction of gender could reveal 'many possible categories embedded in social experiences and social practices' (Lorber, 1996, p. 143), as gender can be considered as both a personal and cultural construction (Chodorow, 1995). Sociologists provide several methodologies for the multidimensional categorization of gender in research designs (Lorber, 1996). Future research can complement this study by framing the concept of gender as multidimensional, rather than polarized, categories and investigate students' perceptions involved in group work processes and peer evaluations in more nuanced detail.

Third, we have to acknowledge that, although we have included some qualitative data, our primary analysis is quantitative in nature. This highlights a more general issue: the analysis of group work processes may be best accomplished with a true mixed methods design that combines quantitative with in-depth qualitative analyses (Creswell & Plano Clark, 2007; Seifert *et al.*, 2010). The latter approach would generate more detailed and nuanced insights into group work processes and students' perspectives, which future research should strive for.

Conclusion

This study investigated the effects of gender on students' cooperative behaviours and performance in group work through the results of self- and peer-assessment of 1001 students. Findings suggest the gender balance in group composition leads to a group work process with less social loafing behaviours and equitable contributions from the group members. However, their performance was found to be lower, and the frequent absence of leaders in gender balanced groups seemed to be one possible reason for this phenomenon. Although the effects could not be determined to be significant, the findings also indicated lower performance of all-male groups, as well as more troublesome group process for all-male groups and male gender exception groups. The students' comments accompanying self- and peer-assessment revealed the nature of group leaders as helpful and creative, whereas the lowly-rated members—more com-

monly male students—were portrayed as being uncommitted, representing the social loafing problem. The overall results underscore the importance of providing students with opportunities of cooperative learning in gender diverse groups where possible.

Acknowledgements

The authors would like to thank Mark Ridolfo and Barras Stone for their help in data collection and constructive comments on earlier drafts.

NOTES

¹ A figure containing the box plots will be provided upon request to the authors.

² We fully acknowledge that the qualitative analysis we provide here is designed to support the main quantitative analysis and as such can only be preliminary in nature. More detailed investigations should follow.

³ The rating of each student's contribution to the group work is calculated with the difference between the average of the ratings each student received and the average of group peer-assessment ratings. The contributions of male students in male gender exception groups are evaluated by their peers, on average, to be as low as -1.044 . This is a notably lower evaluation compared to the evaluation for the male students in gender balanced groups (mean = -0.095) or all male students (mean = -0.070). Although there is a tendency of the same pattern observed with the female gender exception groups, the evaluation received by female students in female gender exception groups was only slightly lower (mean = -0.035) than the ratings female students in gender balanced groups received (mean = 0.101) or the mean rating of all female students (0.148).

References

- Adams, R. B. & Ferreira, D. (2009) Women in the boardroom and their impact on governance and performance, *Journal of Financial Economics*, 94(2), 291–309.
- Bacon, D. R., Stewart, K. A. & Anderson, E. S. (2001) Methods of assigning players to teams: A review and novel approach, *Simulation & Gaming*, 32(1), 6–17.
- Bacon, D. R., Stewart, K. A. & Silver, W. S. (1999) Lessons from the best and worst student team experiences: How a teacher can make the difference, *Journal of Management Education*, 23(5), 467–488.
- 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(5), 663–680.
- Baker, D. F. (2008) Peer assessment in small groups: A comparison of methods, *Journal of Management Education*, 32(2), 183–209.
- Bernard, B. (1997) Girls' world or anxious times: What's really happening at school in the gender war?, *Educational Review*, 49(3), 221–228.
- Bolton, M. K. (1999) The role of coaching in student teams: A 'just-in-time' approach to learning, *Journal of Management Education*, 23(3), 233–250.
- Bonanno, H., Jones, J. & English, L. (1998) Improving group satisfaction: Making groups work in a first-year undergraduate course, *Teaching in Higher Education*, 3(3), 365–382.
- Boud, D. (2001) Introduction: Making the move to peer learning, in: D. Boud, R. Cohen, J. Sampson (Eds) *Peer learning in higher education: Learning from and with each other* (London, Routledge), 1–20.
- Boud, D., Cohen, R. & Sampson, J. (1999) Peer learning and assessment, *Assessment & Evaluation in Higher Education*, 24(4), 413–426.
- Boud, D., Cohen, R. & Sampson, J. (Eds) (2001) *Peer learning in higher education: Learning from and with each other* (London, Routledge).
- 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(5), 268–272.

- Budaev, S. V. (1999) Sex differences in the Big Five personality factors: Testing an evolutionary hypothesis, *Personality and Individual Differences*, 26(5), 801–813.
- Burke, R. J. (1969) Some preliminary data on the use of self-evaluations and peer ratings in assigning university course grades, *Journal of Educational Research*, 62(10), 444–448.
- Carli, L. L. (2001) Gender and social influence, *Journal of social Issues*, 57(4), 725–741.
- Carpenter, M. A. (2002) The implications of strategy and social context for the relationship between top management team heterogeneity and firm performance, *Strategic Management Journal*, 23(3), 275–284.
- Carpenter, M. A., Geletkanycz, M. A. & Sanders, W. G. (2004) Upper echelons research revisited: Antecedents, elements, and consequences of top management team composition, *Journal of Management*, 30(6), 749–778.
- 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(2), 117–127.
- Chapman, K. J., Meuter, M., Toy, D. & Wright, L. (2006) Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes, *Journal of Management Education*, 30(4), 557–569.
- Cheng, W. & Warren, M. (2000) Making a difference: Using peers to assess individual students' contributions to a group project, *Teaching in Higher Education*, 5(2), 243–255.
- Chodorow, N. J. (1995) Gender as a personal and cultural construction, *Signs: Journal of Women in Culture and Society*, 20(3), 516–544.
- Ciani, K. D., Summers, J. J., Easter, M. A. & Sheldon, K. M. (2008) Collaborative learning and positive experiences: Does letting students choose their own groups matter?, *Educational Psychology*, 28(6), 627–641.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd edn) (Hillsdale, NJ, Lawrence Erlbaum).
- Connerley, M. L. & Mael, F. A. (2001) The importance and invasiveness of student team selection criteria, *Journal of Management Education*, 25(5), 471–494.
- Costa, P. T. Jr, Terracciano, A. & McCraem, R. R. (2001) Gender differences in personality traits across cultures: Robust and surprising findings, *Journal of Personality and Social Psychology*, 81(2), 322–331.
- Cotton, K. (2001) *Developing employability skills*. Available online at: http://educationnorthwest.org/webfm_send/524 (accessed 18 July 2010).
- Craig, J. M. & Sherif, C. W. (1986) The effectiveness of men and women in problem-solving groups as a function of group gender composition, *Sex Roles*, 14(7/8), 453–466.
- Creswell, J. W. & Plano Clark, V. L. (2007) *Designing and conducting mixed methods research* (Thousand Oaks, CA, Sage).
- Decker, R. (1995) Management team formation for large-scale simulations, in: J. D. Overby, A. L. Patz (Eds) *Developments in business simulation and experiential exercises* (Statesboro, GA, Association for Business Simulation and Experiential Learning).
- Druskat, V. U. & Wolff, S. B. (1999) Effects and timing of developmental peer appraisals in self-managing work groups, *Journal of Applied Psychology*, 84(1), 58–74.
- Eagly, A. H. (1987) *Sex differences in social behavior: A social-role interpretation* (Hillsdale, New Jersey, Erlbaum).
- Earley, P. C. (1989) Social loafing and collectivism: A comparison of the United States and the People's Republic of China, *Administrative Science Quarterly*, 34(4), 565–581.
- Ellis, P. D. (2010) *The essential guide to effects sizes* (Cambridge, Cambridge University Press).
- Elwood, J. (2005) Gender and achievement: What have exams got to do with it? *Oxford Review of Education*, 31(3), 373–393.
- Elwood, J. & Murphy, P. (2002) Tests, tiers and achievement: Gender and performance at 16 and 14 in England, *European Journal of Education*, 37(4), 398–416.
- Fall, R., Webb, N. M. & Chudowsky, N. (2000) Group discussion and large-scale language arts assessment: Effects on students' comprehension, *American Educational Research Journal*, 37(4), 911–941.

- Feichtner, S. B. & Davis, E. A. (1984) Why some groups fail: A survey of students' experiences with learning groups, *Journal of Management Education*, 9(4), 58–73.
- Feingold, A. (1994) Gender differences in personality: A meta-analysis, *Psychological Bulletin*, 116(3), 429–456.
- Fellenz, M. (2006) Toward fairness in assessing student group work: A protocol for peer evaluation of individual contributions, *Journal of Management Education*, 30(4), 570–591.
- Finkelstein, S., Hambrick, D. C. & Cannella, A. A. J. (2009) *Strategic leadership—theory and research on executives, top management teams, and boards* (Oxford, Oxford University Press).
- Freeman, K. A. (1996) Attitudes toward work in project groups as predictors of academic performance, *Small Group Research*, 27(2), 265–282.
- Gammie, E. & Matson, M. (2007) Group assessment at final degree level: An evaluation, *Accounting Education: an international journal*, 16(2), 185–206.
- Graen, G. B., Hui, C. & Taylor, E. A. (2006) Experience-based learning about LMX leadership and fairness in project teams: A dyadic directional approach, *Academy of Management Learning & Education*, 5(4), 448–460.
- Grissom, R. J. (1994) Probability of the superior outcome of one treatment over another, *Journal of Applied Psychology*, 79(2), 314–316.
- Hedges, L. V. (1981) Distribution theory for Glass's estimator of effect size and related estimators, *Journal of Educational Statistics*, 6(2), 106–128.
- Hernandez, S. A. (2002) Team learning in a marketing principles course: Cooperative structures that facilitate active learning and higher level thinking, *Journal of Marketing Education*, 24(1), 73–85.
- Jehn, K. A., Northcraft, G. B. & Neale, M. A. (1999) Why differences make a difference: A field study of diversity, conflict, and performance in workgroups, *Administrative Science Quarterly*, 44(4), 741–763.
- Jenkins, M. & Johnson, G. (1997) Linking managerial cognition and organizational performance: A preliminary investigation using causal maps, *British Journal of Management*, 8(s1), 77–90.
- Johnson, C. B. & Smith, F. I. (1997) Assessment of a complex peer evaluation instrument for team learning and group processes, *Accounting Education (JAI)*, 2(1), 21–40.
- Johnson, D. W. & Johnson, R. T. (1999) Making cooperative learning work, *Theory into Practice*, 38(2), 67–73.
- Johnston, L. & Miles, L. (2004) Assessing contributions to group assignments, *Assessment & Evaluation in Higher Education*, 29(6), 751–768.
- Joshi, A., Liao, H. & Roh, H. (2011) Bridging domains in workplace demography research: A review and reconceptualization, *Journal of Management*, 37(2), 521–552.
- Kaenzig, R., Hyatt, E. & Anderson, S. (2007) Gender differences in college of business educational experiences, *Journal of Education for Business*, 83(2), 95–100.
- Katzenbach, J. R. (1997) The myth of the top management team, *Harvard Business Review*, 75(6), 83–91.
- Knaak, S. (2004) On the reconceptualizing of gender: Implications for research design, *Sociological Inquiry*, 74(3), 302–317.
- Lejk, M. & Wyvill, M. (2001) 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(6), 551–561.
- Lipsey, M. W. & Wilson, D. B. (2001) *Practical meta-analysis* (Thousand Oaks, Sage Publications).
- Liu, N. & Carless, D. (2006) Peer feedback: The learning element of peer assessment, *Teaching in Higher Education*, 11(3), 279–290.
- Lorber, J. (1993) Believing is seeing: Biology as ideology, *Gender & Society*, 7(4), 568–581.
- Lorber, J. (1996) Beyond the binaries: Depolarizing the categories of sex, sexuality, and gender, *Sociological Inquiry*, 66(2), 143–159.
- Magin, D. (2001) Reciprocity as a source of bias in multiple peer assessment of group work, *Studies in Higher Education*, 26(1), 53–63.
- Maznevski, M. L. (1994) Understanding our differences: Performance in decision-making groups with diverse members, *Human Relations*, 47(5), 531–552.

- McCorkle, D. E., Reardon, J., Alexander, J. F., Kling, N. D., Harris, R. C. & Iyer, R. V. (1999) Undergraduate marketing students, group projects, and teamwork: The good, the bad, and the ugly?, *Journal of Marketing Education*, 21(2), 106–117.
- McGraw, K. O. & Wong, S. P. (1992) A common language effect size statistic, *Psychological Bulletin*, 111(2), 361–365.
- McLeod, P. L., Lobel, S. A. & Cox, Jr, T. H. (1996) Ethnic diversity and creativity in small groups, *Small Group Research*, 27(2), 248–264.
- Mello, J. A. (1993) Improving individual member accountability in small work group settings, *Journal of Management Education*, 17(2), 253–259.
- Miller, T. & Triana, M. D. (2009) Demographic diversity in the boardroom: Mediators of the board diversity-firm performance relationship, *Journal of Management Studies*, 46(5), 755–786.
- Milliken, F. J. & Martins, L. L. (1996) Searching for common threads: Understanding the multiple effects of diversity in organizational groups, *Academy of Management Review*, 21(2), 402–433.
- Murphy, P. & Elwood, J. (1988) Gendered experience, choices and achievement—exploring the links, *International Journal of Inclusive Education*, 2(2), 95–118.
- Nicol, D. (2010a) From monologue to dialogue: Improving written feedback processes in mass higher education, *Assessment & Evaluation in Higher Education*, 35(5), 501–517.
- Nicol, D. (2010b) *The foundation for graduate attributes: Developing self-regulation through self and peer assessment*. Available online at: <http://qmwww.enhancementthemes.ac.uk/docs/publications/the-foundation-for-graduate-attributes-developing-self-regulation-through-self-and-peer-assessment.pdf> (accessed 24 May 2012).
- Nicol, D. J. & Macfarlane-Dick, D. (2006) Formative assessment and self-regulated learning: A model and seven principles of good feedback practice, *Studies in Higher Education*, 31(2), 199–218.
- Nielsen, S. (2009) Why do top management teams look the way they do? A multilevel exploration of the antecedents of TMT heterogeneity, *Strategic Organization*, 7(3), 277–305.
- Nielsen, S. (2010) Top management team diversity: A review of theories and methodologies, *International Journal of Management Reviews*, 12(3), 301–316.
- Orlitzky, M. & Benjamin, J. D. (2003) The effects of sex composition on small-group performance in a business school case competition, *Academy of Management Learning and Education*, 2(2), 128–138.
- Paulus, T. M., Bichelmeyer, B., Malopinsky, L., Pereira, M. & Rastogi, P. (2005) Power distance and group dynamics of an international project team: A case study, *Teaching in Higher Education*, 10(1), 43–55.
- Pfaff, E. & Huddleston, P. (2003) Does it matter if I hate teamwork? What impacts student attitudes toward teamwork, *Journal of Marketing Education*, 25(1), 37–45.
- Pitt, M. J. (2000) The application of games theory to group project assessment, *Teaching in Higher Education*, 5(2), 233–241.
- Robbins, T. L. & Fredendall, L. D. (2001) Correlates of team success in higher education, *Journal of Social Psychology*, 141(1), 135–136.
- Ryan, G. J., Marshall, L. L., Porter, K. & Jia, H. (2007) Peer, professor and self-evaluation of class participation, *Active Learning in Higher Education*, 8(1), 49–61.
- Sackett, P. R., DuBois, C. L. & Noe, A. W. (1991) Tokenism in performance evaluation: The effects of work group representation on male–female and White–Black differences in performance ratings, *Journal of Applied Psychology*, 76(2), 263–267.
- Sadler, D. R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18(2), 119–144.
- Sadler, D. R. (2010) Beyond feedback: Developing student capability in complex appraisal, *Assessment & Evaluation in Higher Education*, 35(5), 535–550.
- Seifert, T. A., Goodman, K., King Patricia, M. & Magolda, M. B. B. (2010) Using mixed methods to study first-year college impact on liberal arts learning outcomes, *Journal of Mixed Methods Research*, 4(3), 248–267.
- Sormunen-Jones, C., Chalupa, M. R. & Charles, T. A. (2000) The dynamics of gender impact on group achievement, *Delta Pi Epsilon Journal*, 42(3), 154–170.

- Stanley, T. D. (2001) Wheat from chaff: Meta-analysis as quantitative literature review, *Journal of Economic Perspectives*, 15(3), 131–150.
- Stark, E. M., Shaw, J. D. & Duffy, M. K. (2007) Preference for group work, winning orientation, and social loafing behavior in groups, *Group & Organization Management*, 32(6), 699–723.
- Taras, M. (2007) Assessment for learning: Understanding theory to improve practice, *Journal of Further and Higher Education*, 31(4), 363–371.
- Tarricone, P. & Luca, J. (2002) Employees, teamwork and social interdependence—a formula for successful business?, *Team Performance Management: An International Journal*, 8(3/4), 54–59.
- Thomas, D. R. (2006) A general inductive approach for analyzing qualitative evaluation data, *American Journal of Evaluation*, 27(2), 237–246.
- Tinklin, T. (2003) Gender differences and high attainment, *British Educational Research Journal*, 29(3), 307–325.
- Topping, K. J., Samuels, J. & Paul, T. (2008) Independent reading: The relationship of challenge, non-fiction and gender to achievement, *British Educational Research Journal*, 34(4), 505–524.
- Tsui, A. S., Egan, T. D. & O'Reilly, C. A. (1992) Being different: Relational demography and organizational attachment. *Administrative Science Quarterly*, 37(4), 549–579.
- Want, K., Lam, Y. R. & Ho, L. (2002) The effects of schooling on gender differences, *British Educational Research Journal*, 28(6), 827–843.
- Warin, J. & Dempster, S. (2007) The salience of gender during the transition to higher education: male students' accounts of performed and authentic identities, *British Educational Research Journal*, 33(6), 887–903.
- Waring, S. M. (2010) The social studies methods course: What do teacher candidates know and want to know about teaching social studies?, *Educational Research & Evaluation*, 16(5), 437–449.
- Warrington, M., Younger, M. & Williams, J. (2000) Student attitudes, image and the gender gap, *British Educational Research Journal*, 26(3), 393–407.
- Watson, W.E., Kumar, K. & Michaelsen, L. K. (1993) Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups, *Academy of Management Journal*, 36(3), 590–602.
- Williams, H. M. & Meân, L. J. (2004) Measuring gender composition in work groups: A comparison of existing methods, *Organizational Research Methods*, 7(4), 456–474.
- Wood, W. (1987) Meta-analytic review of sex differences in group performance, *Psychological Bulletin*, 102(1), 53–71.
- Woodfield, R., Jessop, D. & McMillan, L. (2006) Gender differences in undergraduate attendance rates, *Studies in Higher Education*, 31(1), 1–22.
- Young, I. M. (1994) Gender as seriality: Thinking about women as a social collective, *Signs: Journal of Women in Culture and Society*, 19(3), 713–738.