4th International Conference on
TEACHING, LEARNING and EDUCATION
10-12 September, 2021 Zurich, Switzerland

To Be or not to Be Part of a Team: Decision Dynamics, Winners and Losers in Teamwork

Stephan Schmucker[1*] and Sönke Häseler[2]

[1]*Lecturer, Faculty of Business, Economics and Social Science, University of Hamburg, Germany*

[2]*Associate Fellow, Competence Centre for a Sustainable University, University of Hamburg, Germany*

Abstract

This paper empirically revisits the question of team performance versus individual performance while taking a closer look at the decision dynamics within the teams. The findings are based on a sample of 108 undergraduate students. Having completed an individual assignment, the students were grouped in dyads and then given the same assignment once again as a team. That way, we were able to trace the team output back to the individual members' input. We find that on average, the teams only marginally outperformed the student in each team who achieved the higher individual result. Moreover, the teams failed to effectively exploit available information about their individual members' relative strength. An analysis of who prevails in intra-team discussions showed that higher-performing students are more likely to assert themselves; however, in many cases they were also mislead by their partner's deviating opinion. Overall, we identify a worrisome amount of 'negative learning', i.e., instances of students emerging from the team exercise knowing less than they did before. The results add to the existing body of evidence that suggests caution when assigning to teams a task that could also be accomplished individually. Our novel analysis of the decision dynamics within the dyads more specifically sheds some light on whether teamwork is beneficial for low- versus high-performance individuals, with some implications for instructors and managers.

Keywords: dyads, team performance, decision dynamics, peer instruction, higher education

1. Introduction

Teamwork is ubiquitous in life. Yet often it is not inevitable. With many tasks, a manager or educator may decide whether to assign the job to an individual or a team. Currently, the pressure to work from home due to the COVID-19 pandemic may provide additional impetus to reconsider the situations in which teamwork is truly necessary. Similarly, workers and students can sometimes freely decide whether they want to work on their own or as part of a team. How should they choose? Clearly, the optimal decision for each person involved will depend on many factors, including the roles available to them, the nature of the task, relative skill levels, etc.

info@ictle.com
www.ictle.com

This paper aims to contribute to the vast literature on the merits of teamwork by focusing on a specific setting. We analyse the results of an experiment among undergraduate students enrolled in a course on Human Resource Management at the University of Hamburg. The students were given a curriculum-related task which they completed twice – first individually and then jointly with a partner, forming a dyad, the smallest possible team. This set-up allows us to address two sets of research questions.

The first one concerns individual versus team performance: Do the dyads on average outperform the individuals, as should be expected? How many individuals performed better alone than they subsequently did with a partner? Does it make a difference for the dyads' performance to know which one of their members did better individually? Secondly, to enrich the lessons that our findings may yield for students, we also investigate decision dynamics: Do the chances of a student prevailing in the dyad's decision-making process depend on their personal characteristics, on their individual performance, or on their being informed how they performed in relation to their partner individually? How often is a correct individual response abandoned in favour of a false team response? The answers have some bearing on the students' decision whether to engage in teamwork and what to expect once they are part of the team. Moreover, the discussion will offer some thoughts for educators on whether to offer voluntary or mandatory teamwork in the first place.

The remainder of this paper is structured as follows. The next section summarises some of the literature that applies to various aspects of our study. Section 3 sketches the research design. Section 4 presents our results pertaining to performance levels (4.1) and decision dynamics (4.2). Finally, the Conclusion offers a brief discussion of the implications that the results may hold for the parties involved in teamwork in higher education and beyond. We also mention some limitations and avenues for future research.

2. Related Literature

The research presented in this paper is part of a series of experiments with students that share the same basic design but differ in focus. The first of the resulting publications, Schmucker & Häseler (2016), describes the methodology of measuring student performance using classroom response systems and relating it to the students' socio-demographic characteristics. Two follow-up publications focus on the students' choice of a team partner and on various potential socio-demographic determinants of team and individual performance. Each paper discusses the relevant theoretical and empirical literature on the interplay between socio-demographics and the formation and performance of teams, and these discussions need not be repeated here. Schmucker & Häseler (2017) specifically look at diversity within the teams, while Schmucker et al. (2019) investigate migration background and social class. Each of these studies has shown that a sizable number of students individually outperformed the teams they later found themselves in. In other words, their performance was dragged down by their partners; they would have been better off on their own, strictly in terms of performance, barring other potential benefits of teamwork, such as the potential for peer instruction (Mazur, 2013). Intuitively, a team should perform no worse than its strongest member working by herself, so this persistent finding came as some surprise. It also begged the question whether the number of cases in which the teams underperformed their individual members could be

reduced if the teams are explicitly told which member achieved the higher individual score. This is one of the questions we now address with our focus on decision dynamics.

The huge body of prior literature on team performance and individual performance, of which we can only highlight some selected examples, unsurprisingly reports mixed results. Unlike with individuals, effectiveness in teams requires smooth coordination processes, which team mental models and transactive memory can facilitate (Kozlowski & Ilgen, 2006). Team mental models refer to the shared, organized understanding and mental representation of beliefs regarding the team's task environment (Cannon-Bowers et al. 1993). They are associated with significant effects on team performance (Kneisel, 2020). Transactive memory in turn refers to the mutual awareness of "who knows what" in a team (Wegner, 1995). The members exchange and update information about everyone's specific knowledge to create a distributed memory system which can be accessed by all members, potentially raising team performance (Zhou & Pazos, 2020). Also, team intelligence is widely assumed to exceed the combined intelligence of the individuals concerned (Hackman, 1987).

On the other hand, decades of group research (Hill, 1982) have shown that groups usually fall short of reasonable productivity baselines. Many empirical studies have supported the concept of process loss in groups as stated in Steiner's (1972) theory of group productivity. Process losses prevent the groups from realizing their full productivity potential. Other studies have used the performance of the most capable team member as a productivity baseline. Contrary to the intuition stated above, in relatively few studies did the teams meet or exceed this baseline (e.g., Laughlin et al., 2003). In addition, O'Boyle and Aguinis (2012) find that individual performance follows a Paretian (power law) distribution, rather than being normally distributed, as often assumed in Human Resource Management. In our context, a Paretian distribution implies that some individuals are all the more likely to exceed team performance. Furthermore, ability disparity, whose extent depends on the distribution of individual performance, may be assumed to affect team performance (Tu et al., 2020). More generally, the exact distribution of individual performance needs to be considered in team management, including team selection and performance management.

Cross et al. (2016) argue that excessive teamwork can lead to collaborative overload, exhausting team members and reducing productivity. In most cases, 20% to 35% of important collaborative work comes from only 3% to 5% of the individuals. Li et al. (2015) concentrate on employees who frequently contribute beyond the scope of their roles – so-called 'extra-milers'. A single extra-miler in a key position can do more for team performance than all the other members combined. In sum, this body of research casts doubt on the intuition that team performance principally exceeds the individual performance of the team members.

Our suspicion that teamwork may yield a lot of false answers in situations where some of the individuals involved would have responded correctly is supported by the so-called Asch effect. It refers to the phenomenon of group consensus and social pressure that induces an individual to change a correct answer in reaction to group members' false answers. Asch (1956) investigated how individuals conform to the opinions of a peer group. In a series of experiments, he found that individuals will often deviate from what they consider to be the correct answer merely in order 'to go along with the group'. The larger the majority group (the number of allies), the more people adapt. This would suggest that the Asch effect is less

pronounced in our work with dyads. Yet, individuals experience stronger emotions in dyads than in (larger) groups, as there is only one relationship through which social emotions can flow (Moreland, 2010).

Our second set of research questions more specifically asks which one of the two team members is more likely to prevail in the discussions. Do the students' personal characteristics play a role? The literature suggests that they do. Ma (2005) explores the relationship between the Big Five personality factors and negotiation styles. He finds that extraversion is associated with confrontational conflict styles while, conversely, agreeableness usually goes along with non-confrontational styles, such as compromising. Extraversion is furthermore associated with interpersonal assertiveness, confidence, and gregariousness (Costa & McCrae, 1995). Anderson & Kilduff (2009) find that more assertive individuals tend to dominate groups because they behave in ways that make them appear more competent than they are. This suggests that such persons may exert more influence than their actual competence warrants, and highly competent group members who are low in trait dominance may be unjustly overlooked. When it comes to predicting which team member asserts herself among dissenting views in the team, the literature thus provides good reason to believe that a team member's personality-related characteristics may have more sway than the actual merits of her opinion, or in short: personality trumps performance.

3. Research Design

The experiment took place as part of a regular lecture on human resource management for first- and second-year undergraduate students at the University of Hamburg. The students had been advised beforehand that an experiment was going to be conducted on that day. Since participation in the lectures is voluntary, so was participation in the experiment. The students were also told that the event was going to test their knowledge of the material taught in previous lectures and that it was to provide the basis for some social science research with direct relevance to the course. Furthermore, the students were asked to arrive for class in pairs. Those who failed to do so were arbitrarily assigned to other individual arrivals. We obtained 54 dyads of students, each of which received a pair of clicker devices by which they were to convey their answers to questions presented on slides.

The experiment proper began with a set of ten questions relating to previous lectures. For each question, five answer choices were presented, only one of which was correct. The students were to respond using their clicker devices but without consulting their partner. The time allowed to read the questions and answers and to submit the response was 90 seconds per question. The number of correct responses out of the ten questions is our measure of individual performance. Next, each student was asked to answer 12 questions pertaining to their socio-demographic background. Then each dyad surrendered one of their two clickers, and a random subset of 28 dyads were informed which member achieved the higher score in the first round of ten questions (in three such dyads, the first-round performances were equal).

Since most of our dyads formed voluntarily while some were composed externally, and since research has suggested that these two modes of team composition may affect team performance and other outcomes (Chapman et al., 2006), the students were additionally asked to indicate how familiar they were with their partner. However, as with most of our previous

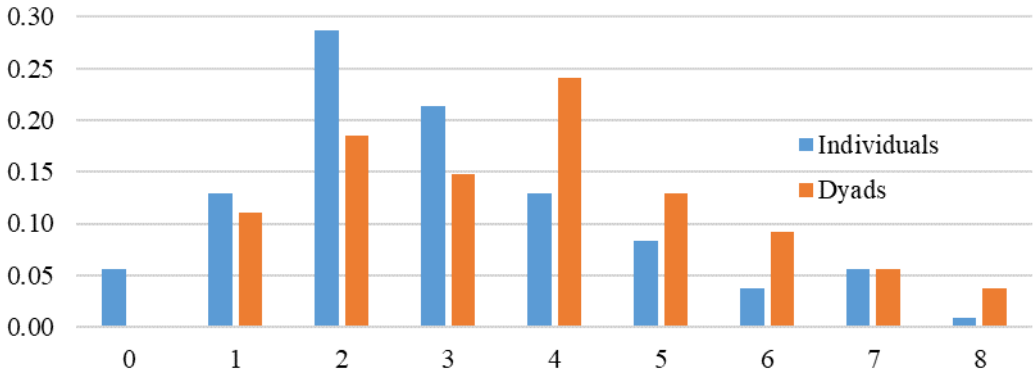
work in similar settings, this control variable made no difference to the results reported below and is therefore not considered any further. Finally, all dyads were again shown the same ten questions about the course contents, though this time a joint answer was to be submitted following some discussion within the dyad (peer instruction, cf. Mazur, 2013). The number of correct responses in this second round of ten questions is our indicator of dyad performance.

4. Results

4.1 Performance Levels

Figure 1 shows the relative frequencies of the number of correct responses for the individual students and for the dyads. Note that with five answer choices in each of ten questions, plain guesswork would have yielded two correct responses. Two is also the most common value for the individual students, though the average is at 2.963. Thus, the task was clearly challenging. The mode for the dyads is four, with an average of 3.778. Some of this performance premium will merely be due to the dyads having had more time to ponder the same set of questions once again.

Figure 1: Relative frequencies of the number of correct responses



The trouble for teamwork, as it were, starts if we consider the maximum individual scores within the dyads. Those students who individually performed at least as well as their partners on average scored 3.833 correct answers – slightly more than the teams, even though the individuals had only one opportunity to consider each question, rather than two. The difference increases further if we only consider those students who strictly outperformed their partners; in other words, dyads with equally performing students are removed from the average. The students with the strictly higher individual results averaged 3.898 correct responses, whereas the corresponding subset of the teams only achieved 3.816 on average. Also, 24 of the 58 teams achieved fewer correct answers than the better ones of their members. In four cases, the team result even fell short of the individual scores of both members – which clearly should not happen. These numbers provide some first indication that, at least in terms of performance, above-average students have little to gain from teamwork. Despite the additional available time, the discussions and decision-making process with their less knowledgeable partners left them worse off than they were on their own.

Similar findings arose in our earlier studies, and a possible explanation could be that unless the members of a team knew each other quite well, they could not know which one of them was more knowledgeable and should therefore have a bigger say in the discussions. If that were the case, we should find that those teams whom we informed that their members had diverging individual performances – i.e., the 28 teams whom we spoke to about individual performance, less the three teams with equal individual scores – achieved higher results than those who did not receive such information. Table 1 presents some information to that effect. The 25 dyads who were told which one of their members did strictly better individually outperformed the 29 'uniformed' dyads, but by much less than a tenth of a correct answer. This difference is far from significant ($t < 0.1$). Regardless of the information status (and in line with the overall average for all dyads, see above), both 'types' of teams achieve fewer correct answers than they would have if they had simply settled for the set of answers provided by their higher-scoring member in the first round ('average max individual performance'). This difference is especially pronounced for the 'informed' dyads, i.e., those who knew whose opinion was more valuable – a counterintuitive result. The more knowledgeable partners allowed their performance to be dragged down, even though everyone in these teams knew that these individuals had a better chance of getting the answers right. In other words, telling the teams whose opinion will likely make a greater contribution to joint performance did not help them make the most of their potential.

Table 1: Descriptive statistics for informed versus uninformed dyads

	Was the dyad informed of a difference in individual performance?	
	no	yes
number of dyads	29	25
average dyad performance	3.759	3.8
average max individual performance	3.586	4.12

What the above univariate analysis does not take into account is the possibility that informing some of the dyads does in fact make a difference to performance, but the two sets of dyads also differ in other performance-related respects whose influence masks that of the performance information. In other words, a more substantial examination requires multivariate analysis. We therefore ran several regressions to explain dyad performance, using as independent variables alternatively the sum and the maximum of the individual performances, as well as all the socio-demographic information we have on the students. Each time, a dummy variable for the informed teams was included as a potential predictor. The results are not reported in detail because they are easily summarised: All the regressions explain dyad performance well but informing half of the dyads never has a significant impact on their performance.

Why is it that the dyads seem not to have exploited the information on relative individual performance that we gave them and which we considered to be valuable for dyad performance? We can only speculate. The first potential explanation is that the information was in fact no news to them. Most students have a fairly accurate idea regarding their own performance, at least in relation to the class average. So even if they knew nothing about their teammate's strength and therefore had to assume that it was average, they should have an idea as to whose opinion should carry more weight. Yet most students did know something about

their partners. Most dyads had formed voluntarily ahead of the experiment, and it seems likely that the students would pick partners they knew. When asked how familiar the team members were with each other, the average response was 1.57 on a scale of 0 to 3 – between "somewhat" and "quite" familiar. This is where the above-mentioned transactive memory may come into play. The concept refers to a system comprised of the individual memories of the members of a group, augmented by the shared memory regarding who knows what (Wegner, 1995). Familiarity, frequent face-to-face communication, and shared experiences foster the formation of transactive memory systems (Lewis, 2004). Our hunch that the information was no news to the students is confirmed by anecdotal evidence: When we informed half of the teams as to who performed better individually, very few students seemed surprised by the announcement. Furthermore, if relative knowledge levels played any role in the team discussions, the 'uninformed' dyads, too, will have found ways to communicate that information.

The other potential reason for the irrelevance of the information that we gave the students is precisely that: The teams did not consider it to be sufficiently relevant. Perhaps maximising dyad performance by exploiting that information was not worth sacrificing other objectives that the team members may have had, be it to maintain an equal say in the decisions, to be polite to each other, to foster a relationship outside of the experiment, or any number of other non-performance considerations, as suggested by 'groupthink' (Janis, 1972).

4.2 Decision Dynamics

By decision dynamics we mean the question as to what drives team decisions in cases where the members submitted different individual responses, i.e., when a decision regarding dissenting views had to be taken. For that purpose, we pick one member of each dyad at random (call her 'A') and create the variable WIN. It signifies how many times over the course of the ten questions A succeeded in having her individual answer from the first round accepted as the team answer in the second round, while her partner's ('B') individual response differed. Instances where A 'lost out' to B in the discussion, i.e., when the team answer equals B's but not A's, make a negative contribution to WIN. The actual values of WIN fall in the range of -8 to +6. The correctness of the answers is initially not considered here. Furthermore, we defined the variable INFO to assume the value of 1 if A was told that she outperformed B, a value of -1 if B did better than A, and a value of 0 if either A's individual score exactly equalled B's or if A was not informed of her relative performance. The socio-demographic variables and the individual performance levels ("INDI") were redefined in terms of differences, i.e., A's value less B's.

Table 2 shows the results of regressing WIN first on INFO, then on INDI, and finally on both variables. According to Model 1, being informed that she outperformed B in the first round of questions significantly raises A's likelihood of asserting herself in the team discussions. In Model 2, the relative individual performance has a similar effect, but the model's explanatory power is twice as high. Model 3 includes both predictors, which are naturally highly correlated (r=0.72). Once INDI is included in the regression, INFO becomes irrelevant, as its coefficient is virtually zero and the R^2 value of Model 3 equals that of Model 2. This suggests that the students' relative performance indeed has a strong influence on the outcome of the discussions within the dyad, and the students do not need to be told which one of them did

better individually – somehow, they just know. Note, however, that our result regarding the relative explanatory power of INFO versus INDI may to some extent be due to the way in which these variables were defined: INFO only assumes three different values and equals zero for half of the sample, whereas INDI varies across all dyads. Therefore, INDI has a better chance of yielding predictive power – not for contentual reasons but simply due to the definition of the variables.

Table 2: Explaining WIN by INFO and INDI

	Model 1	Model 2	Model 3
INFO	1.821*** (0.497)		-0.022 (0.626)
INDI		0.78*** (0.132)	0.785*** (0.192)
observations	54	54	54
R^2	0.205	0.402	0.402
Constant term not reported. Standard errors in parentheses. (***) statistically significant at the 1% level.			

As a robustness check, we also ran regressions that additionally included all the socio-demographic information we have on the students. Selected specifications are reported in Table 3, which for comparability mirrors Table 2 in terms of the usage of INFO and INDI. Two points are worth noting about the results. Firstly, what we find concerning INFO and INDI almost exactly reflects the results we obtained in Table 2 without the socio-demographic variables: In Model 4, INFO makes a significant contribution to explaining who wins in the team discussions, and so does INDI in Model 5. However, as above, when both indicators are included in the regression (Model 6), we find that the influence of INFO almost completely vanishes, and the variable makes no contribution to R^2 as we move from Model 5 to Model 6. Again, we interpret this to mean that the dyads allocate a greater say in the team decision to the better-performing member, but they do not need to be told who that person is.

Table 3: Explaining WIN by INFO, INDI, and the socio-demographic indicators

	Model 4	Model 5	Model 6
INFO	1.83** (0.83)		0.088 (1.044)
INDI		0.693*** (0.196)	0.676** (0.29)
FEMALE	-1.665* (0.8)	-1.235* (0.642)	-1.26 (0.725)
AGE	-2.175*** (0.657)	-1.884*** (0.563)	-1.892*** (0.59)
LOAN	2.524** (0.982)	2.035** (0.856)	2.028** (0.889)
PARENTS	-0.321* (0.168)	-0.288* (0.143)	-0.29* (0.148)
TRAINING	4.162*** (1.039)	3.524*** (0.912)	3.535*** (0.952)
ACCESS	-1.718 (1.217)	-1.491 (1.022)	-1.475 (1.075)
CITIZEN	-11.766*** (3.966)	-11.146*** (3.317)	-11.201*** (3.495)
LANGUAGE	12.001*** 3.844	10.541*** (3.32)	10.547*** (3.436)
REGION	5.768** (2.536)	6.337** (2.158)	6.32** (2.242)
observations	26	26	26
R^2	0.703	0.786	0.786
Constant term not reported. Standard errors in parentheses. ***, ** and * refer to statistical significance at the 1%, 5% and 10% level.			

Secondly, we find significant effects on 'winning' in the team discussions for a range of socio-demographic variables. However, beyond noting that a person's personal background clearly affects their degree of assertiveness and persuasiveness within a team, we refrain from interpreting the individual coefficient estimates. First, this is because of the small sample size – the students failed to provide answers to all the questions, and reformulating the variables in terms of intra-dyad differences exacerbated the problem of missing values. Second, some of the indicators are quite highly correlated among each other, so there is some multicollinearity.

In a second set of analyses regarding decision dynamics, we now turn to the correctness of the responses. The discussions within the dyads will often help the students improve their knowledge – they learn from their partners about the correct answer, where previously they could only guess or had mistaken beliefs: peer instruction at work. However, the reverse may also occur, and these are the cases we are particularly interested in: Sometimes students abandon their correct individual answer in favour of an incorrect team answer, yielding to the (mistaken) arguments of their partner. In such instances, we might speak of mislearning, or of a 'learning fail': good knowledge has been replaced with bad. Two types of such 'learning fails' can be distinguished. A 'minor fail' is the case where the team response is incorrect while exactly one individual response was correct. A 'major fail' occurs if the team response is wrong while both individual responses were correct.

Arguably, the occurrence of such fails should depend on whether the teams were informed of the partners' relative performance: Regarding minor fails, an 'informed' team knows which partner is more likely to be correct with respect to any given question, so this person's opinion should carry more weight in the discussions, leading to fewer fails. Regarding major fails, an 'informed' team has less need for discussions, and the discussions appear to have contributed to the fail, since both partners were correct before them. Table 4 presents some statistics to this effect. With 54 teams, ten questions, and 39 instances of missing answers, we are left with 501 complete decision situations – 501 sets of two individual responses and one

team decision. Overall, we count 68 'minor fails' and 39 'major fails'. Therefore, in more than 21% of the decision situations, at least one partner came out of the discussions 'knowing' less than they did going in – in a sense the downside of peer instruction and teamwork.

Table 4: Instances of 'mislearning'

	all dyads	Was the dyad informed of a difference in individual performance?	
		no	yes
all dyad decisions	501	261	240
'minor fails'	68 (13.57%)	33 (12.64%)	35 (14.58%)
'major fails'	39 (7.78%)	18 (6.9%)	21 (8.75%)

Our hypothesis on the effect of informing the teams of the members' relative performance is not borne out by the data: Both types of 'fails' occur more frequently among informed than among uninformed teams. Thus, knowing about relative performance levels within the team does not help to mitigate mislearning, one of the downsides of teamwork and peer instruction.

5. Conclusion

5.1 Summary and Discussion

The effectiveness of teams as opposed to individuals has attracted some 70 years of academic research (Kozlowski & Ilgen, 2006). For a small contribution to this vast literature, we have sought to illuminate a particular finding from our earlier research, which is that individuals outperform teams of two surprisingly often in direct comparison. Does it make a difference for the dyads to know which one of their members had a stronger individual performance? If so, what might be the finer dynamics behind the teams' decision-making?

We had the members of 54 dyads of undergraduate students respond to a set of ten course-related questions, first individually and then again as a team, after consulting with their partner. Half of the teams were informed as to whether who, if anyone, of their members performed better individually in the first round of questions. With 3.778 versus 2.963 correct responses on average, the teams clearly outperformed the individuals. However, this may in part be due to the teams simply having extra time to consider the same questions once more. More interestingly, those individuals who performed at least as well individually as their partners averaged 3.883 correct answers in the first round – more than the teams, despite the individuals' disadvantage in terms of decision time. Those who did strictly better than their partners even averaged 3.898. This is already worrisome from the perspective of teamwork.

It could be argued that, especially with a task like ours, teamwork cannot have much benefit unless the partners know each other quite well and therefore have an idea as to whose opinion should carry more weight in the team's deliberations. Our data show the performance differential between the 'informed' and the 'uninformed' teams to be far below a tenth of a correct answer. And even – indeed, especially – in the 'informed' teams, dyad performance falls considerably short of the average score achieved by the better-performing members. These individuals allowed their performance to be dragged down in teamwork, even though they and their partners both knew about the relative individual performance. None of these results materially change in multivariate analysis, i.e., when controlling for the socio-demographic information we had about the students. We can only assume that personality

factors, which our analysis does not account for, play a role in the team discussions. In particular, it seems likely that the influence of trait dominance (Anderson & Kilduff, 2009) outweighed actual expertise in the discussions, to the detriment of team performance. Groupthink (Janis, 1972) may provide another explanation, suggesting that the pursuit of harmony in the dyads is more important than asserting oneself over one's partner. Finally, social loafing may also explain the teams' failure to make the most of their situation. Latané et al. (1979) found that participants working in teams exert less effort than participants who worked by themselves. A lack of incentives can further reduce team effort. In our experiment, there were no extrinsic incentives whatsoever, and anonymity precluded any opportunity for the teams to distinguish themselves (e.g., Miles & Greenberg, 1993).

For a more detailed picture, we also looked at who prevailed in the team discussions. We find that the rate of a student 'winning' over her partner depends more on her relative individual performance than on her being informed that she outperformed her partner. So again, the information about the members' relative individual performance made little difference to team outcomes. These results also hold in multivariate analysis. Looking at individual decisions furthermore allowed us to gauge the extent of 'mislearning', i.e., instances where, rather than learning the correct response from their partner as 'peer instruction' promises, students abandon their correct individual answer in favour of a wrong team answer. We find that in 21% of the decision situations, at least one partner emerged from the discussions 'knowing' less than they did going in. Moreover, these 'learning fails' actually occur more frequently in the teams that were informed of the relative individual performance – not less frequently, as one might have expected and, indeed, hoped for the sake of effective teamwork.

Conscious of the limitations of our study, we daresay that the results do not bode well for the effectiveness of teamwork, at least in our very specific context: Despite more favourable circumstances, the dyads are often trumped by their better-performing members. Moreover, it appears that the teams do not use the available information on relative individual performance effectively – the consequence being that many of the students would be better off on their own. A more nuanced interpretation of the results regarding the merits of teamwork very much depends on the party concerned: A *weak student* unambiguously stands to profit from teamwork – in all likelihood, her partner will be more knowledgeable than her, so she can both learn from them and enjoy their strong contribution to the team effort. Conversely, *stronger students* run the risk of having their performance diluted by weaker teammates and allowing their good knowledge to be corrupted by their partners' lower-quality input ('negative learning'). Finally, *instructors* should be aware of these redistributive effects. They should keep in mind the tendency, where teamwork is optional, for the stronger students to prefer working by themselves, leaving the weaker students to team up among themselves. Thus, for peer instruction to work, teamwork should be mandatory for all. To the extent that the results can be generalised to the corporate setting, *managers* should consider very carefully whether to prescribe teamwork when the task could also be completed by individuals. The value-added of teamwork is anything but certain, but it certainly comes at a cost. And unless the teams are properly incentivised, maximising their performance can easily become a secondary objective for the team members.

5.2 Limitations and Future Research

Our study is subject to a number of limitations, only some of which can be mentioned here. First, our measure of 'performance' is a very narrow one. The interaction within the teams may have yielded many benefits besides answering the course-related questions – benefits that our measure of performance does not capture and that do not accrue in individual work: learning from each other, getting to know each other, practicing team skills, etc. Similarly, the students may have pursued any number of other objectives besides performing well in the sense defined by us. For example, in the classroom as in any real-life setting, team members can be expected to behave in a way that fails to maximise output in terms of the task but instead promises personal rewards that materialise outside of the situation at hand, such as trading favours with other team members. Thus, our results may be biased in the sense that perhaps the teams were not trying, as hard as the individuals were, to achieve what we were measuring. But then again, this bias is also highly relevant to assessing the relative merits of teamwork in any practical setting. Secondly, the generalisability of the results is doubtful. Dyads are a very special kind of team, and our classroom setting is very different from, say, a team situation in business. Furthermore, the task we posed is not ideally suited to assessing the merits of teamwork. For example, it did not require any creativity – one of the areas where teamwork shines –, nor was there any scope for a division of labour.

Future research might address some of these limitations with an improved experimental setup, e.g., regarding the complexity of the task, or using a longitudinal design. Another interesting avenue for additional research surrounds the question as to why certain team members tend to assert themselves over others. So far, we have only looked at prior individual performance and a small set of socio-demographic factors. Additionally, we might expect personality traits to play a strong role (Ma, 2005; Anderson & Kilduff, 2009). In particular, self-efficacy (Bandura, 1997) – a person's self-assessment as to how well they can cope with contingencies – is likely to affect behaviour within the teams. Confirming such an influence empirically could yield important messages for prospective team members.

Acknowledgment

The authors thank the Equal Opportunity Fund of the University of Hamburg and the eLearning Office of the University's Faculty of Business, Economics and Social Sciences for support. Ingo Häseler and Manuel Harder contributed valuable ideas.

References

Anderson, C. and Kilduff, G.J. (2009), "Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance", *Journal of Personality and Social Psychology*, Vol. 96 No. 2, pp.491-503.

Asch, S.E. (1956), "Studies of independence and conformity: A minority of one against a unanimous majority", *Psychological Monographs*, Vol. 70 No. 9, pp.1-70.

Bandura, A. (1997), *Self-efficacy: The exercise of control*, Freeman, New York, NY.

Cannon-Bowers, J.A., Salas, E. and Converse, S.A. (1993), "Shared mental models in expert team decision making", Castellan, N.J. (Ed.), *Individual and group decision making: Current issues*, Erlbaum, Hillsdale, NJ, pp. 221-246.

Chapman, K.J., Meuter, M., Toy, D. and 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*, Vol. 30 No. 4, pp.557-569.

Costa, P.T. and McCrae, R.R. (1995), "Solid ground in the wetland: A reply to Block", *Psychological Bulletin*, Vol. 117, pp.216-220.

Cross, R., Rebele, R. and Grant, A. (2016), "Collaborative Overload", *Harvard Business Review*, Vol. 1, pp.74-79.

Hackman, J.R. (1987), "The design of work teams", Lorsch, J.W. (Ed.), *Handbook of Organizational Behavior*, Prentice-Hall, Englewood Cliffs, pp.315-324.

Hill, G.W. (1982), "Group versus individual performance: Are N+1 heads better than one?", *Psychological Bulletin*, Vol. 91 No. 3, pp.517-539.

Janis, I. (1972), *Victims of groupthink: A psychological study of foreign-policy decisions and fiascos*, Houghton Mifflin, Boston.

Kneisel, E. (2020), "Team reflections, team mental models and team performance over time", *Team Performance Management*, Vol. 26 No. 1/2, pp. 143-168.

Kozlowski, S.W.J. and Ilgen, D.R. (2006), "Enhancing the Effectiveness of Work Groups and Teams", *Psychological Science in the Public Interest*, Vol. 7 No. 3, pp.77-124.

Latané, B., Williams, K. and Harkins, S. (1979), "Many hands make light the work: The causes and consequences of social loafing", *Journal of Personality and Social Psychology*, Vo. 37 No. 6, pp.822-832.

Laughlin, P.R., Gonzalez, C.M. and Sommer, D. (2003), "Quantity estimations by groups and individuals: Effects of known domain boundaries", *Group Dynamics: Theory, Research, and Practice*, Vol. 7 No. 1, pp.55-63.

Lewis, K. (2004), "Knowledge and performance in knowledge-worker teams: A longitudinal study of transactive memory systems", *Management Science*, Vol. 50, pp.1519-1533.

Li, N., Zhao, H.H., Walter, S.L., Zhang, X.-A. and Yu, J. (2015), "Achieving more with less: Extra milers' behavioral influences in teams", *Journal of Applied Psychology*, Vol. 100 No. 4, pp.1025-1039.

Ma, Z. (2005), "Exploring the Relationships between the Big Five Personality Factors, Conflict Styles, and Bargaining Behaviors", *IACM 18th Annual Conference*.

Mazur, E. (2013), *Peer Instruction: A User's Manual*. Kindle Edition, Pearson, Boston, MA.

Miles, J.A. and Greenberg, J. (1993), "Using punishment threats to attenuate social loafing effects among swimmers", *Organizational Behavior and Human Decision Processes*, Vol. 56 No. 2, pp.246-265.

Moreland, R.L. (2010), "Are dyads really groups?", *Small Group Research*, Vol 41 No 2, pp.251-267.

O'Boyle Jr., E. and Aguinis, H. (2012), "The best and the rest: Revisiting the norm of normality of individual performance", *Personnel Psychology*, Vol. 65, pp.79-119.

Schmucker, S. and Häseler, S. (2016), "Measuring the impact of student diversity on performance with classroom response systems", *The Online Journal of Distance Education and e-Learning*, Vol. 4 No. 4, pp.15-22.

Schmucker, S. and Häseler, S. (2017), "Formation, diversity, and performance of student teams of two: Some experimental evidence", *Zeitschrift für Diversitätsforschung und -management*, Vol. 2 No. 2, pp.92-105.

Schmucker, S., Häseler, S. and Sprengel, J. (2019), "Team Selection and Performance: The Role of Migratory Background and Social Class", *The European Proceedings of Social & Behavioural Sciences*, pp.255-270.

Steiner, I.D. (1972), *Group Process and Productivity*. Academic Press, New York, NY.

Tu, Y., Hong, Y., Jiang, Y. and Zhang, W. (2020), "Team ability disparity and goal interdependence influence team members' affective and informational states", *Group Dynamics: Theory, Research, and Practice, Vol. 24 No* 1, pp.6-25.

Wegner, D. M. (1995), "A computer network model of human transactive memory", *Social Cognition*, Vol. 13, pp.319-339.

Zhou, Z. and Pazos, P. (2020), "Empirical perspectives of transactive memory systems: a meta-analysis", *Team Performance Management*, Vol. 26 No. 7/8, pp. 409-427.