

Shall We Explain? Augmenting Learning from Intelligent Tutoring Systems and Peer Collaboration

Robert G.M. Hausmann, Brett van de Sande, and Kurt VanLehn

Pittsburgh Science of Learning Center, University of Pittsburgh,
3939 O'Hara Street, Pittsburgh, Pa, 15260-5179
{bobhaus,bvds}@pitt.edu, vanlehn@cs.pitt.edu

Abstract. Learning outcomes from intelligent tutoring systems (ITSs) tend to be quite strong, usually in the neighborhood of one standard deviation. However, most ITS designers use the learning outcomes from expert human tutoring as the gold standard (i.e., two standard deviations). What can be done, with the current state of the art, to increase learning from an ITS? One method is to modify the learning situation by asking students to use the ITS in pairs. To enhance performance, we drew upon the beneficial effects of structured peer collaboration. The results suggest that the intervention was successful. Pairs of students solved more problems and requested fewer bottom-out hints than individuals. To test the possibility that the effect was due to the best partner in the group directing the problem solving, a nominal groups analysis was conducted. A nominal group is a statistical pairing of the non-interacting individuals' performance. The results from the nominal groups replicated the same pattern of results, but with a reduced magnitude. This suggests that the best member may have contributed to some of the overall success of the pair, but does not completely explain their performance.

Keywords: Collaborative learning; explanation activities; studying examples.

1 Introduction

An often-heard suggestion is that students may learn more from an intelligent tutoring system (ITS) if two students worked together on the system instead of working on it alone. Early studies did not support this hypothesis, and instead suggested that having pairs of students using an ITS produced the same learning gains as having students work alone as they used it [early studies with Lisp & geometry tutors]. However, null results are often open to many interpretations, such as a lack of statistical power. This issue has been re-opened recently, and for good reasons. The simplest reason for re-opening the pair/solo hypothesis is that, despite the designers' best efforts, the hints given by an ITS are sometimes more confusing than helpful. Perhaps having two students interpret the hints may help alleviate this problem.

Another reason for studying pairs using an ITS is that learning gains can be impressive when the students in a pair actually collaborate, as opposed to one student dominating the problem solving. Collaborators can encourage each other's reasoning, notice and productively resolve conflicts in their thinking, confirm each other's beliefs, externalize

their thoughts, and so on. Collaborators could use an ITS to catch mistakes that the students manage to overlook, to provide hints when they have exhausted their mutual knowledge, or to resolve conflicts that they cannot resolve themselves. In short, meta-cognitive strategies exist for using an ITS as an effective scaffolding for peer problem solving.

Lastly, many observers have noticed that students working alone often abuse the ITS by asking for more help or less help than they need [1]. In particular, many ITSs give increasingly specific hints when asked, and the last “bottom-out” hint tells the student exactly what step to enter. Some students ask for bottom-out hints on almost every step. Conversely, some students will enter incorrect versions of a step dozens of times without asking for a hint. Although it is possible that students do not know how to use the help system effectively, experiments with a help-seeking tutor have shown that explicitly teaching students effective help-seeking strategies did not change their long-term behavior [2]. They went back to abusing the ITS as soon as the help-seeking tutor was replaced by the regular ITS. This suggests that students know how to seek help effectively, but they sometimes choose otherwise.

One way to reduce the frequency of help misuse may be to have students work in pairs. If both students know that rapidly pressing the hint button in order to get the bottom-out hint is bad for learning, then such abuse would only occur if they both simultaneously agree to be “bad.” Similarly, they would have to both agree to enter mistaken entries repeatedly without asking for help, even though they both know this could be a waste of time. So perhaps having students work in pairs on an ITS could increase the frequency of proper help usage, compared to students working alone.

In short, there are at least three reasons why pairs of student should learn more than individuals using an ITS: (1) pairs may be able to interpret the ITS’s hints more successfully, (2) the ITS may help student collaborate effectively, and (3) pairs of students are less likely to abuse the ITS than individuals.

Although the first and third hypotheses are somewhat novel, there has been a considerable amount of work on the second hypothesis [3, 4]. For example, Rummel and Spada [5] contrasted learning and problem solving under four different conditions. The first was a detailed, worked-out example of successful collaboration. Participants were asked to study the example of smooth collaboration, and apply it to their own dialog. A second group was provided with a collaboration script, in which case their interactions were structured in a way that was hypothesize to promote successful collaborative outcomes. Finally, there were two control conditions that did not structure the collaborative interactions. They found that the example and scripted conditions demonstrated better understanding of successful collaboration principles, as well as a better understanding of the domain (i.e., therapy and diagnosis of medical and psychological disorders) than the control conditions.

The laboratory study by Rummel and Spada suggests that successful collaborative interactions can be instructed and scripted. Walker et al. [6] extended this finding by developing a cognitive tutor that teaches students how to interact in the context of peer tutoring. They developed a peer-tutoring script that assists students along three dimensions. The first dimension prepares students for peer tutoring by providing instruction on the domain content, as well as relevant pedagogical strategies. The second dimension involves actual tutoring, whereby the tutor is taught to set goals for the student, as well as monitor the student’s progress during problem solving. The last

dimension introduces skills for effective interaction, such as providing elaborated help. These three dimensions were used to construct a cognitive tutoring system for peer collaboration, thus reifying the empirical results on effective peer tutoring and collaboration.

Although the available evidence suggests that ITSs can have a positive impact on collaborative learning (i.e., hypothesis 2), research addressing the collaborative use of ITS hints has been relatively sparse (i.e., hypotheses 1 and 3). However, the empirical work on collaboratively studying worked-out examples, which is reviewed in the next section, may be of some relevance because requesting a string of bottom-out hints can turn a problem into a worked-out example.

1.1 Collaborative Example Studying

So far, we have discussed the hypothesis that pairs would be more effective than individuals when they work on an ITS, but similar remarks apply to studying examples as well. By *example*, we mean a problem plus a presentation of the multiple steps required for its solution. When individuals study an example, they sometimes self-explain it by filling in the gaps between steps, relating the inter-step reasoning to prior knowledge, etc. [7]. Sometimes students do anticipatory self-explanation, where they try to generate the next solution step themselves, then look at the example to see if they are right [8]. Prompting can increase the amount of self-explanation [9, 10]. However, even with prompting, students can abuse an example just as they can abuse an ITS. They abuse an example by simply reading it shallowly and not trying to self-explain much of it.

When discussing pairs using an ITS, we suggested that they may be more effective than individuals for 3 reasons: (1) pairs may be able to interpret the ITS's hints more successfully, (2) the ITS may help student collaborate effectively, and (3) pairs of students are less likely to abuse the ITS than individuals. Those same three reasons apply to examples as well. (1) A pair may be more able to interpret an examples' reasoning more successfully than an individual student. (2) An example scaffolds collaborators by helping them extend their reasoning when they get stuck, resolve conflicts productively, externalize their reasoning, confirm their beliefs, etc. (3) Pairs are less likely to abuse an example than individuals.

Laboratory evidence, which suggests that pairs may be better suited for studying examples than individuals, can be found in [11]. In their experiment, participants were asked to study some instructional materials collaboratively, then solve LISP programming problems individually. The pairs' performance on the LISP problems was contrasted with individuals studying the instructional materials alone. They found that the programming performance of the pairs was significantly better than the solo student performance; however, the authors note that the advantage for collaboration diminished over time.

The present study is another step toward understanding if and when "two heads are better than one" for learning. The study compares pairs vs. solo students who are both studying examples and solving problems with an ITS. As they study examples, they are prompted to self-explain. This study is preliminary in that we did not use pre-tests and post-tests, and thus cannot measure students' learning gains. However, we did record their behavior during the training in order to determine if it was affected by the solo/pair manipulation. Thus, this study counts as a *manipulation check* for a subsequent study of learning gains.

Although verbal protocols were collected, they have not yet been fully analyzed, so this paper reports only the analyses of log files generated by the tutoring system. Using them, we found evidence that pairs abused the help system less frequently than solo students, as predicted. We also looked for but failed to find signs of collaborative facilitation, in that pairs would make fewer errors due to collaboration than nominal pairs. On the other hand, the pairs did no worse than the nominal pairs, so there was no process loss [12]. Thus, all the current evidence is positive—two heads may indeed be better than one for explaining examples and solving problems with an ITS.

1.2 Problem Solving and Example Studying in an ITS

The Andes physics tutor was initially developed as a replacement for paper and pencil homework problems. The advantage for solving problems with Andes is the adaptive support it provides to the student. One form of adaptive support is the on-demand hints, which are provided in a graded fashion. Typically, the first hint points the student's attention to a relevant feature of the problem (i.e., a *Pointing Hint*). The second hint level presents general instructional principles that are relevant to the problem (i.e., a *Teaching Hint*). Finally, at the terminal level, the bottom-out hint tells the student exactly which action to take (i.e., a *Bottom-out Hint*).

In addition to the on-demand hints, Andes provides a series of videos that the students can watch in order to learn about various solution strategies, as well as how to use elements of the Andes interface. For the purposes of the current study, we modified the available videos so that they were broken down into individual problem-solving steps. Typically, students are responsible for starting and stopping the videos, but that generally leads to shallow cognitive processing of the video content. Instead, at the juncture of each step, we prompted the students to engage in an explanatory activity. Solo students and pairs were prompted to generate explanations while studying video-based, worked-out examples. The purpose of prompting students was to increase their cognitive processing of the examples.

The use of examples in the present experiment slightly diverges from traditional studies in the sense that students were prompted to engage in explanation while studying an isomorphic worked-out example *after* solving a related problem. We used this design for two reasons. First, the ACT-R theory of learning suggests that students only learn from the correct application of knowledge [13]. Second, the cognitive load associated with problem solving can impede a deep encoding of the problem-solving goals and operators [14].

2 Method

The following experiment was designed to test the effects of collaboration on problem solving and example studying while using an ITS, primarily with an emphasis on the collaborative use of hints.

2.1 Participants

Thirty-nine undergraduates, enrolled in a second semester physics course, were randomly assigned to one of two experimental conditions: solo students ($n = 11$) or pairs

($n = 14$). Volunteers were recruited from several sections of a second-semester physics course, which covered Electricity and Magnetism. Participants were recruited during the third week of the semester, with the intention that the experimental materials would coincide with their introduction in the actual physics course. The participants were paid \$10 per hour. To ensure that the participants' motivation remained high during the entire two-hour session, they were offered an incentive of an additional \$10 for doing well on the tests, which they all received.

2.2 Materials

The materials developed for this experiment were adapted from an earlier experiment [15]. The domain selected for this experiment was electro-dynamics with a focus on the definition of the electric field, which is expressed by the vector equation: $\mathbf{F} = q\mathbf{E}$. This particular topic is typically covered within the first few weeks of a second-semester physics course. Thus, it is an important concept for students to learn because it represents their first exposure to the idea that a field can exert a force on a body.

To instruct the participants, several materials were developed. Four electro-dynamics problems were created. These problems are representative of typical problems found at the end of a chapter in a traditional physics textbook. The problems covered a variety of topics, including the definition of the electric field, Newton's first and second law, the weight law, and several kinematics equations. Each of the four problems was implemented in Andes. Andes was chosen because its design allowed for both the presentation of video-based examples, as well as coached problem solving. The first problem served as a warm-up problem because none of the students had any prior experience with the Andes user interface.

In addition to the problems, three examples were created in collaboration with two physics instructors at the U.S. Naval Academy. The examples contained a voice-over narration of an expert solving the problems, and they were structured such that they were isomorphic to the immediately preceding problem.

2.3 Procedure

The procedure consisted of several activities. The first activity was to watch a short, introductory video on the Andes user interface. Afterwards, the participants read instructions on how to produce explanations, including an example. Next, participants were asked to use Andes to solve a warm-up problem. The experimenter was available to answer any user-interface questions. He was not, however, allowed to give away any domain-specific information. During problem solving, the student had access to the flag feedback, the hint sequences, and an Equation Cheat Sheet. Once the student submitted a final answer, she then watched and explained an example of an expert solution of an isomorphic problem. The example solutions were broken down into steps, and at the conclusion of each step the student was prompted to explain (either individually or collaboratively). Once the explanation was complete, the participant clicked a button to go onto the next step. Only the cover story and given values differed between the problem-solving and example problems. The students alternated between solving problems and studying examples until all four problems were solved and all three examples were studied, or until two hours elapsed.

2.4 Measures

Several dependent measures, taken from the log files, were used to assess problem-solving performance, including: the number of entries, correct entries, solution rate, and the number of bottom-out hint requests.

3 Results

The results are organized into two sections. The first reports performance differences between the solo students and pairs at the problem level. The second section then reports the same dependent measures using a “nominal group analysis,” which is considered the gold standard for collaborative research [16]. A nominal group is a statistical pairing of the non-interacting individuals. To construct a nominal group, individuals from the solo condition were randomly paired, and the best performance from each individual was taken to represent the pair.

3.1 Performance Differences

Before delving into the problem-solving performance measures, we first analyzed the solution rates (i.e., whether or not a final answer was found) for two reasons. First, the students worked at their own pace; and second, the experiment was capped at two hours. Thus, there was no guarantee that all of the students would finish all of the problems in the allotted time. The pairs were much more likely to submit an answer to the final problem than the solo students ($\chi^2 = 4.81, p = .03$).

An analysis of the mean number of entries and correct entries for the final problem confirmed the solution rate results. The pairs ($M = 34.29, SD = 6.72$) demonstrated more entries than the solos ($M = 25.00, SD = 14.97$), $F(1, 23) = 4.32, p = .05, d = .87$. Moreover, the pairs ($M = 23.29, SD = 5.06$) demonstrated reliably more correct entries for the final problem than the solos ($M = 15.64, SD = 9.85$), $F(1, 23) = 6.36, p < .02, d = 1.06$. Taken together, these results suggest that the pairs were more efficient in solving the problems during the two-hour experiment.

To test if the participants abused the available help, bottom-out hint requests were analyzed. Requesting multiple bottom-out hints is an indication that the student required more direct instruction, and this may have translated into *gaming the system* behaviors. However, if the student requests a reasonable number of bottom-out hints, then that is an indication that she is more interested in connecting the information found in the instructional materials (i.e., examples) to the individual problem-solving steps.

The bottom-out hint usage interacted with time, such that the difference between conditions was apparent early in the experiment, but diminished over time. To correct for Type II errors due to multiple statistical tests across the four problems, a repeated measures ANOVA was used. The univariate tests, which contrasted the two conditions for each problem, indicated that the solos demonstrated marginally higher bottom-out hint requests for the warm-up problem ($F(1, 21) = 3.98, p = .06$), reliably more hint requests for the first problem ($F(1, 21) = 7.64, p = .01$), and no reliable differences for the final two problems (see Fig. 1).

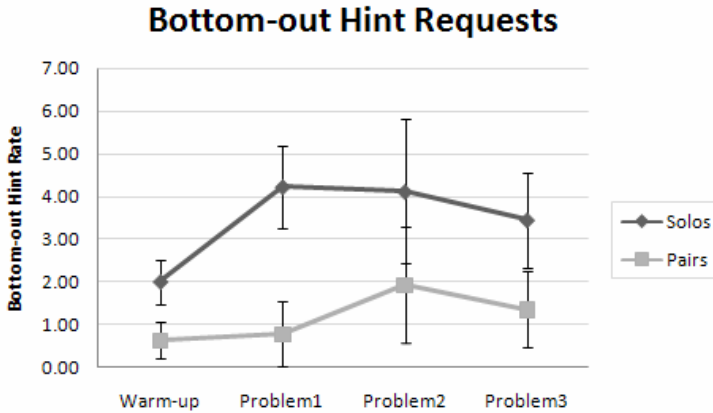


Fig. 1. The mean number of bottom-out hint requests per problem for solos and pairs

3.2 Nominal Group Analyses

One of the dangers of working in a collaborative setting is the threat of requiring more time to complete a task than it would when working alone. To test if there was a loss in efficiency, we compared real pairs to the nominal pairs by measuring the amount of time taken to input each correct entry. The results suggest there were no time penalties for working in a group. In fact, there was a small amount of evidence to the contrary. On average, real pairs ($M = 47.96, SD = 17.17$) demonstrated faster times between correct entries for the first problem than the nominal pairs ($M = 67.05, SD = 16.45$), $F(1, 23) = 7.89, p = .01, d = 1.18$.

In addition, the dependent measures used to contrast problem-solving performance between the two experimental conditions were repeated for the real pairs and the

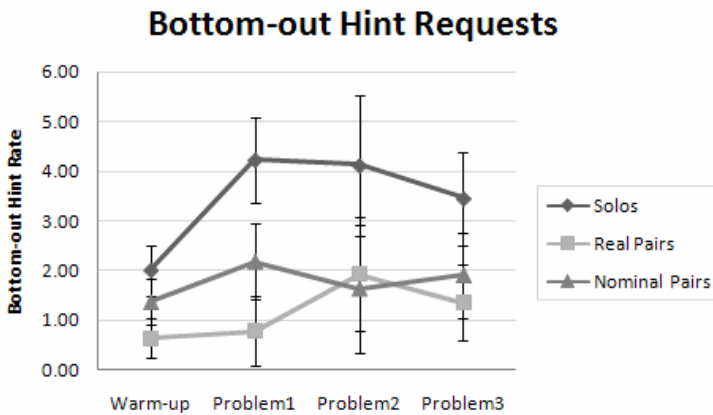


Fig. 2. The mean number of bottom-out hint requests per problem, including nominal pairs

nominal pairs. For the final problem, the nominal pairs ($M = 33.54$, $SD = 9.15$) submitted an equal number of entries for the final problem as the real pairs ($M = 34.29$, $SD = 6.72$). However, the number of correct entries made for the final problem was replicated. There was a marginal difference between the two groups in terms of the correct number of entries made on the last problem, $F(1, 23) = 3.16$, $p = .09$, $d = .75$. The real pairs ($M = 23.29$, $SD = 5.06$) entered marginally more correct entries for the final problem than the nominal pairs ($M = 18.73$, $SD = 7.73$).

In terms of the bottom-out hint requests, the pattern of results was also consistent with the solo results (see Fig. 2). The real pairs still requested fewer bottom-out hints for the Problem 1 than the nominal pairs, $F(1, 23) = 5.11$, $p = .03$, $d = .95$. None of the other contrasts reached traditional levels of statistical significance.

4 Discussion

The introduction to this paper proposed three hypotheses regarding collaboration during solving problems with an ITS. The first hypothesis stated that pairs may be in a better position to profit from an ITS's hints than individuals because each student may interpret the hint in a different way. Through the process of sharing and debugging their various interpretations, pairs of students can help each other make sense of the hints. Evidence for this claim can be found in both the completion rate and the use of bottom-out hints. The pairs progressed further into the problem set than the individuals, and they required fewer bottom-out hints to finish the problems.

The second hypothesis stated that a step-based ITS may help students collaborate more effectively. Although, the present study did not directly test this hypothesis (i.e. by contrasting the frequency of successful collaborative processes for step-based tutoring with a more open learning environment), we indirectly tested the hypothesis by conducting a *nominal groups analysis*. Nominal groups were formed by randomly pairing individuals from the solo condition and taking the best performance from each member. For example, if Solo Member A asked for 3 bottom-out hints, and Solo Member B asked for 2, then the score for that nominal group on the bottom-out hint measure was "2." However, if Solo A correctly imputed 8 steps, and Solo B entered 5 correct steps, then the score for that nominal pair was "8." Therefore, the source of the nominal pair's score could come from a different individual for the different measures of problem-solving performance.

The results from the nominal-groups analysis replicated the set of results from the solos. Although the magnitude of the differences between pairs and solos was reduced, the same trend of results was observed. This suggests that there was something special about participating in a collaborative discussion while solving problems with an ITS. That is, the tutoring system helped to scaffold the dialog between interacting students above and beyond the performance of the non-interacting individuals.

Finally, the third hypothesis stated that pairs of students should be less likely to abuse the ITS than individuals because students have a general sense of the proper use of a learning environment. Stated differently, having a partner keeps the individuals honest. Evidence for the third hypothesis was most directly demonstrated with the bottom-out hint requests. Pairs of students requested an average of 67.6% fewer bottom-out hints across the entire 2-hour experiment. The difference in bottom-out hint

requests between the pairs and solos was most pronounced after studying the first example (i.e., Warm-up Problem = 67.9% vs. Problem 1 = 85.6%). This suggests that the pairs may have also been less likely to abuse the examples. Instead of shallowly processing the content, they may have better comprehended and later reused the information in the examples. In the future, we plan to test this hypothesis more directly by analyzing the verbal protocols produced while studying the examples.

In summary, the results from each of the three hypotheses suggest that asking students to solve problems collaboratively, with a step-based tutoring system, is a productive way to enhance learning from an ITS. This study, which served as a positive example of a manipulation check, suggests that future experiments continue to examine the boundary conditions under which collaboration is effective in an ITS. Additional measures of learning need to be used to evaluate the strength of the learning that results from collaboration. For example, the present study does not indicate if learning from collaboration will transfer to individual problem solving and to novel domains. Additional research is needed to answer these and related questions.

Acknowledgements. This work was supported by the Pittsburgh Science of Learning Center, which is funded by the National Science Foundation award number SBE-0354420.

References

1. Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z.: Off-Task behavior in the cognitive tutor classroom: When students game the system. In: Proceedings of ACM CHI 2004: Computer-Human Interaction, pp. 383–390 (2004)
2. Roll, I., Alevan, V., McLaren, B.M., Koedinger, K.R.: Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) *Artificial intelligence in education: Building technology rich learning contexts that work*, vol. 158, pp. 203–210. IOS Press, Amsterdam (2007)
3. Barros, B., Conejo, R., Guzman, E.: Measuring the effect of collaboration in an assessment environment. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) *Artificial intelligence in education: Building technology rich learning contexts that work*, vol. 158, pp. 375–382. IOS Press, Amsterdam (2007)
4. Biswas, G., Leelawong, K., Schwartz, D.L., Vye, N.: The Teachable Agents Group at Vanderbilt: Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence* 19, 363–392 (2005)
5. Rummel, N., Spada, H.: Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences* 14, 201–241 (2005)
6. Walker, E., McLaren, B.M., Rummel, N., Koedinger, K.: Who says three's a crowd? Using a cognitive tutor to support peer tutoring. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) *Artificial intelligence in education: Building technology rich learning contexts that work*, vol. 158, pp. 399–406. IOS Press, Amsterdam (2007)
7. Chi, M.T.H., Bassok, M.: Learning from examples via self-explanations. In: Resnick, L.B. (ed.) *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, pp. 251–282. Lawrence Erlbaum Associates, Inc., Hillsdale (1989)

8. Renkl, A.: Learning from worked-out examples: A study on individual differences. *Cognitive Science* 21, 1–29 (1997)
9. Aleven, V.A.W.M.M., Koedinger, K.R.: An effective metacognitive strategy: Learning by doing and explain with a computer-based Cognitive Tutor. *Cognitive Science* 26, 147–179 (2002)
10. Chi, M.T.H., DeLeeuw, N., Chiu, M.-H., LaVancher, C.: Eliciting self-explanations improves understanding. *Cognitive Science* 18, 439–477 (1994)
11. Bielaczyc, K., Pirolli, P., Brown, A.L.: Collaborative explanations and metacognition: Identifying successful learning activities in the acquisition of cognitive skills. In: Ram, A., Eiselt, K. (eds.) *Proceedings of the Sixteenth Annual Cognitive Science Society Conference*, pp. 39–44. Laurence Earlbaum Associates, Hillsdale (1994)
12. Steiner, I.D.: *Group processes and productivity*. Academic Press, New York (1972)
13. Anderson, J.R., Corbett, A.T., Koedinger, K., Pelletier, R.: Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences* 4, 167–207 (1995)
14. Sweller, J.: The worked example effect and human cognition. *Learning and Instruction* 16, 165–169 (2006)
15. Hausmann, R.G.M., VanLehn, K.: Explaining self-explaining: A contrast between content and generation. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) *Artificial intelligence in education: Building technology rich learning contexts that work*, vol. 158, pp. 417–424. IOS Press, Amsterdam (2007)
16. Lorge, I., Fox, D., Davitz, J., Brenner, M.: A survey of studies contrasting the quality of group performance and individual performance. *Psychological Bulletin* 55, 337–372 (1958)