



<http://www.diva-portal.org>

This is the published version of a paper published in *Educational Psychology*.

Citation for the original published paper (version of record):

Stenlund, T., Jönsson, F., Jonsson, B. (2017)

Group discussions and test-enhanced learning: individual learning outcomes and personality characteristics.

Educational Psychology, 37(2): 145-156

<https://doi.org/10.1080/01443410.2016.1143087>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-117603>

Group discussions and test-enhanced learning: individual learning outcomes and personality characteristics

Tova Stenlund^a, Fredrik U. Jönsson^b and Bert Jonsson^a

^aDepartment of Psychology, Umeå University, Umeå, Sweden; ^bDepartment of Psychology, Stockholm University, Stockholm, Sweden

ABSTRACT

This paper focuses on the factors that are likely to play a role in individual learning outcomes from group discussions, and it includes a comparison featuring test-enhanced learning. A between-groups design ($N = 98$) was used to examine the learning effects of feedback if provided to discussion groups, and to examine whether group discussions benefit learning when compared to test-enhanced learning over time. The results showed that feedback does not seem to have any effect if provided to a discussion group, and that test-enhanced learning leads to better learning than the discussion groups, independent of retention interval. Moreover, we examined whether memory and learning might be influenced by the participants' need for cognition (NFC). The results showed that those scoring high on NFC remembered more than those who scored low. To conclude, testing trumps discussion groups from a learning perspective, and the discussion groups were also the least beneficial learning context for those scoring low on NFC.

ARTICLE HISTORY

Received 5 February 2015
Accepted 13 January 2016

KEYWORDS

Learning; testing effect; cooperative learning; need for cognition

Collaborative or cooperative learning is widely applied in educational settings and it is often seen as a valuable learning condition by educators. An often encountered argument is that a group can achieve more than individuals working on their own (Kirschner, Paas, & Kirschner, 2009b; Slavin, Hurley, & Chamberlain, 2003; Van Blankenstein, Dolmans, Van der Vleuten, & Schmidt, 2013; Vojdanoska, Cranney, & Newell, 2010). A significant part of both collaborative and cooperative learning is learning by participating in group discussions. For both teachers and students, it is important to know which learning strategies are most effective, as well as how individual differences may affect learning. Hence, in the present study, we investigated the learning effects of group discussions with and without feedback on individual performances, and assessed how the effects of learning are related to individual differences with respect to the need for cognition (NFC). NFC is a personality characteristic defined as 'an individual's tendency to engage in and enjoy thinking' (Cacioppo & Petty, 1982, p. 116). Moreover, in both cases, the effects of group discussions on learning are compared with those of test-enhanced learning, an effective learning technique with much scientific support.

Kirschner, Paas, and Kirschner (2009a) defined collaborative learning as 'learning in a group in which knowledge and/or information may be divided across individuals, but where the group as a whole carries out the task' (p. 32). Cooperative learning is defined simply as students working together to achieve shared learning goals (Slavin et al., 2003). In this study, the term *group learning* will be used to represent the concepts of cooperative and collaborative learning. Engaging in reflective activities in

CONTACT Tova Stenlund  tova.stenlund@psy.umu.se

a group, in which individuals verbalise explanations and evaluate problems and solutions, has been shown to be beneficial for learning (Baker & Lund, 1997). Nevertheless, the effects of group learning activities are not clear cut. Research in this area shows contradicting results regarding whether group activity is beneficial for learning (Kester & Paas, 2005; Slavin et al., 2003).

Studies on group learning effectuated in an unconstrained environment show mixed and sometimes negative results regarding learning, while research on group learning implemented in a more constrained environment has demonstrated that this method is positive for the learner (Kirschner et al., 2009b). Thus, more constrained environments, which ensure that the learners engage in more effective group activity, have the potential to aid learners in maintaining learned information (Morgan, Whorton, & Gunsalus, 2000); these environments also stimulate learners to engage in valuable activities, such as verbalising explanations (van Boxtel, van der Linden, & Kanselaar, 2000). Additionally, constrained environments also seem to be more beneficial for developing higher order skills when compared to traditional lecture-based learning. For example, studies comparing groups and individuals show that group learning is more beneficial than individual learning for complex problem-solving tasks, and that individual learning is more beneficial than group learning for more simple recall tasks (Kirschner et al., 2009b). Nevertheless, a constrained environment does not ensure beneficial effects of group learning, as some studies fail to show positive effects even under constrained conditions (see e.g. Mäkitalo, Weinberger, Häkkinen, Järvelä, & Fischer, 2005; Slavin et al., 2003; van Bruggen, Kirschner, & Jochems, 2002). Research in this area also suggests that group learning can have negative effects on retention. Roediger, Meade, and Bergman (2001) argued that the information presented in a group discussion might lead participants in the group to develop false memories, as they may misattribute this information to the original learning context. However, by providing corrective feedback during discussion, one might prevent individuals from acquiring false memories. Indeed, corrective feedback following incorrect answers helps individuals arrive at the correct answer (Vojdanoska et al., 2010), and it seems to be a very effective way of improving learning (Pashler, Cepeda, Wixted, & Rohrer, 2005). In addition to corrective feedback, additional exposure to the material can provide an opportunity to self-correct, without being explicitly corrective.

Group discussions are, as previously indicated, a significant part of group learning and they have been defined as a group of individuals that come together for verbal communication to make decisions or simply share knowledge (Morgan et al., 2000). In an educational context, the teacher often introduces concepts or questions to discuss, or the group analyses a problem or carries out an assigned task. Hence, group discussions are viewed within the context that the learning takes place when completing a well-defined task.

In contrast to group learning, test-enhanced learning (i.e. repeated testing as a no-stakes learning activity) has consistently, and across a wide range of materials, been shown to have beneficial effects on learning, particularly with respect to long-term learning (see, e.g. Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Finn & Roediger, 2013; Jönsson, Hedner, & Olsson, 2012; Kornell, Hays, & Bjork, 2009; Roediger & Karpicke, 2006a, 2006b; Stenlund, Sundström, & Jonsson, 2014). A general finding in this domain is that test-enhanced learning improves an individual's learning more than just repeatedly reading certain material (Roediger & Karpicke, 2006a). This phenomenon is usually referred to as the testing effect, which is often explained by the theory that testing requires active retrieval, whereas reading and re-reading the material simply involve encoding (Karpicke & Roediger, 2008). Research has shown that various factors, such as the number of practice tests, the timing of the test, and test feedback, influence the testing effect. Several studies have found that the effect on learning and memory consistently improves with the number of practice tests (e.g. Karpicke & Roediger, 2007, 2010). Research also shows that spaced learning is more effective than learning grouped in the same or adjacent sessions (e.g. Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Feedback is additionally important because it appears to enhance both short- and long-term learning (Butler, Karpicke, & Roediger, 2008; Kang, McDermott, & Roediger, 2007; Roediger & Butler, 2011). Extensive scientific support related to the testing effect has convinced researchers that individual test-enhanced learning should be implemented in classroom settings to promote learning (Chan, McDermott, & Roediger, 2006; McDaniel, Roediger, &

McDermott, 2007; Roediger & Karpicke, 2006a). However, as was pointed out by Kornell, Rabelo, and Klein (2012), there are only a few studies comparing test-enhanced learning with more active learning approaches, such as group discussions.

Additional and important factors to consider when examining learning effects are individual differences, such as enjoying thinking of and elaborating on questions or problems. Individuals frequently decide for themselves how and what they learn, and it is likely that their learning is related to how they usually process information (Evans, Kirby, & Fabrigar, 2003; Heijne-Penninga, Kuks, Hofman, & Cohen-Schotanus, 2010). One psychological factor related to learning is the NFC (Cacioppo & Petty, 1982). NFC describes variations in motivation and effort in the individuals' cognitive processing (van Seggelen-Damen, 2013). Motivation is also a factor that is particularly important for successful group learning, and it has been examined in a number of studies in this field (Slavin et al., 2003). From this perspective, individuals with a high NFC are more motivated to think about and seek information for sense-making purposes, while individuals with low NFC are more likely to rely on others (e.g. experts) to provide them with this information (Cacioppo, Petty, Feinstein, & Jarvis, 1996). NFC as a concept has been studied by psychologists for many years (Cohen, Stotland, & Wolfe, 1955; Maslow, 1943, etc.), and Cacioppo and Petty (1982) eventually developed a scale to measure NFC. High NFC has been shown to have a beneficial influence on learning, and differences among individuals in their tendencies to engage in and enjoy thinking have been widely studied in a variety of contexts; see Cacioppo et al. (1996) for a review. More recently, a study investigating the types of academic motivations that predict student attendance at a special class for gifted students found that NFC was the strongest predictor (Meier, Vogl, & Preckel, 2014).

In summary, there is mixed evidence as to whether group discussions are beneficial for learning; in addition, only a few studies have investigated individual performance as a function of group discussions. Test-enhanced learning, on the other hand, seems to be an effective individual learning technique. However, test-enhanced learning has rarely been compared to other more meaningful learning strategies (but see Karpicke & Blunt, 2011, for a comparison with concept mapping). For both test-enhanced learning and group discussions, feedback has shown to be effective for correcting and enhancing performance. Finally, a largely ignored aspect of learning is how the characteristics of individual engagement and enjoyment in thinking (i.e. NFC) relate to performance when one is required to participate in group discussions.

Purpose and research questions

The purpose of the current study was to compare a collaborative form of learning – namely, group discussions – with another well-known but individual learning strategy: test-enhanced learning. This study focuses on the individual learning outcomes of group learning; group learning is operationalised as group discussions of a previously studied material. Prior studies have convincingly shown that test-enhanced learning is an effective learning technique (e.g. Dunlosky et al., 2013), but it has, to our knowledge, never been compared with the individual learning outcomes of group discussions. We also examined how the presence or absence of corrective feedback during group discussions affects student learning, with the expectation that discussion followed by feedback should lead to better performance than discussion without feedback (Kelley & McLaughlin, 2012). Thus, test-enhanced learning with feedback is seen as the baseline against which group discussions with and without feedback are compared.

Further, we investigated whether individual differences in NFC affect group discussions. Previous studies have shown that NFC is related to information acquisition, reasoning and problem solving (Cacioppo et al., 1996). Hence, it is reasonable to expect that participants with higher NFC should outperform their lower scoring counterparts. By investigating NFC and its effect on learning in group discussions, and when compared to test-enhanced learning, we intend to shed light on whether individuals with specific characteristics benefit more from one learning strategy over the other. In this respect, we are particularly interested in how NFC relates to the efficacy of group discussions, as it is possible to presume that those scoring low on NFC do not engage as much or as actively in group discussions,

and that this may negatively impact the usefulness of group discussions for the low-scoring group. To measure learning outcomes, we assessed retention individually after 15 min, and again after one and four weeks.

Method

Participants

The participants included 131 students – 106 females (81%) and 25 males – in an upper-secondary school in the northern part of Sweden. The sample was homogenous with respect to ethnicity. The participants' ages ranged from 16 to 19 years, with a mean age of 17.2 years ($SD = .61$). Students who were not present on all four occasions were excluded, and no other exclusion criteria were used; a total of 98 participants completed the study. The mean age in this cohort was 17.2 years ($SD = .59$) and 80% were female. With regard to these background variables, they can be regarded as a representative sample.

Written consent was obtained from all participants. The school board, principals and teachers of the involved institution permitted the study. This study was conducted within the students' original class schedule and was completed in their own classrooms. No financial compensation was given to the participants.

Design

A 3×3 mixed factorial design was used with group serving as the between-subjects factor and retention interval as the within-subjects factor. The participants were randomly assigned to one of three groups, which will be referred to as the discussion without feedback ($n = 35$; 27 females and eight males), discussion with feedback ($n = 31$; 27 females and four males) and test-enhanced learning groups ($n = 32$; 25 females and seven males). To ensure that the groups were comparable, grades (nine-year compulsory school), age and gender were compared. The results of a one-way ANOVA showed that there were no significant differences in grades, $F(2, 95) = .83, p = .439$, or age, $F(2, 95) = .20, p = .822$, between the three groups. Further, a chi-square statistical analysis showed that the three groups did not differ significantly with regards to gender, $\chi^2(2, n = 98) = 1.23, p = .541$. These results suggest successful random group assignment. During the actual discussions, the two discussion groups were further randomly subdivided into small discussion groups (10 groups, respectively) with four to five students in each subgroup. The random assignment led to a configuration where some small discussion groups consisted of females alone, while no groups consisted of solely males. With respect to retention interval, all participants were tested with the same items as during the learning occasions. This was completed 15 min after the last learning session (denoted as the 'immediate test'), one week later, and four weeks later (denoted as the 'One week' and 'Four weeks' retention tests).

Materials

The study materials included 23 pages from a chapter on human emotions from a Swedish psychology textbook (Rydén & Wallroth, 2008), which was divided into 16 short passages covering different affects and their functions.

In total, 16 open-ended questions were developed, one for each of the 16 short passages in the textbook chapter. The 16 questions differed in complexity. Thus, some questions were directed towards factual knowledge and asked for *who*, *what*, or *which*, requiring only a one- or two-word answer. For example, *Who explained primary affects as short and relatively distinctive?* Meanwhile, some questions were directed to more complex knowledge and asked for *why*, *in what way*, or *what does it mean that* – mainly requesting explanations – thus requiring longer answers when compared to the former. An example of this type of question is, *In what way are affects such as disgust important to us?* Even though

the questions differ somewhat in complexity, the internal consistency for all questions was .83 (which represented the mean coefficient alpha for the three retention tests), suggesting content homogeneity.

Instruments

Questionnaire with background information and experience of the text

A questionnaire was administered asking about the participants' age and gender, and three questions about their experience with the textbook material. First, the participants were asked how much of the text they had read; 63% of the participants read the entire textbook material during the study period, while another 33% had read more than half of the text. Second, the participants were asked how much of the information in the textbook material was familiar to them; 11% answered that most of the text was familiar, 38% answered that about half the information in the text was familiar, while 44% found either a small part or nothing (7%) to be familiar. Third, the participants were asked how they perceived the textbook material. Overall, 25% perceived the text as easy or very easy, 42% perceived the text as neither easy nor difficult, and 30% considered the text to be somewhat difficult. Chi-squared statistical analysis showed that the three groups did not differ significantly with regards to how much of the text they had read, $\chi^2(4, n = 98) = 6.47, p = .17$; how much of the text was familiar to them, $\chi^2(6, n = 98) = 2.87, p = .82$; or how difficult the text was perceived to be, $\chi^2(6, n = 98) = 10.72, p = .09$.

Questionnaire to determine how the discussion groups spent their time

A questionnaire was also administered to the discussion groups to examine how much time they spent discussing the different questions. Twenty-eight per cent responded that they used about half of the time to discuss the questions. About 40% responded that they used more than half of the time or nearly all of the time to discuss the questions, and about 32% reported that they used less than half of the time or hardly any time to discuss the questions. The students were also asked if they devoted more time to specific questions. Seventy-three per cent responded that they devoted more time to the more demanding questions, 16.5% spent equal amounts of time to the two types of questions, and the remainder reported that they did not know. It is important to note that the discussion groups without feedback and the discussion groups with feedback did not differ significantly in the amount of time spent to discuss the different questions, $\chi^2(2, n = 66) = 5.48, p = .06$.

Mental Effort Tolerance Questionnaire – NFC

NFC was measured using the short version of the *Mental Effort Tolerance Questionnaire* (METQ; Dornic, Ekehammar, & Laaksonen, 1991), which is a Swedish adaptation of Cacioppo and Petty's (1982) original *Need for Cognition Scale*. The questionnaire consists of 30 items rated on a five-point Likert scale (1 = strongly disagree; 3 = neutral; 5 = strongly agree), with a possible score range from 30 to 150. Twelve of the statements represented positive attitudes toward engaging in and enjoying thinking, while 18 indicated negative attitudes. The items indicating negative attitudes were reversed before calculating the METQ score; therefore, high scores were associated with high NFC. In Dornic et al. (1991), the scale was considered to be valid, with one clear dominant factor (eigenvalue = 9.12) – a result that is in line with the original NFC scale (Cacioppo et al., 1996); this scale was also reliable (coefficient $\alpha = .90$).

Procedure

Data were collected in the fall semester of 2012. On day one, all participants were given 40 min to read the chapter about human emotions (the textbook material), after which the questionnaire, with questions related to background information, as well as questions pertaining to how the students experienced the text was administered. The study phase was followed by a practice phase in which the participants participated in either the test-enhanced learning group or one of the two discussion groups.

In the first practice session, the test-enhanced learning group repeated the test two times (approximately 20 min per test). The order of the items was randomised for each test. The items were displayed

using a Microsoft PowerPoint slideshow; the students recorded their responses using an answer sheet (pen and paper). A slide showing the correct answer (feedback) was given before the next question. During the same amount of time (40 min), the participants in the discussion groups with and without feedback discussed the same 16 questions with the textbook material in hand. At the end of the session, the participants in the discussion group with feedback were provided with feedback (i.e. the correct answer to the questions).

On the following day, a second practice session took place. Similar to the first practice session, the participants in the test-enhanced learning group completed two practice tests and feedback was provided after each question; the discussion groups discussed the same questions with the textbook material at hand, and the correct answer was provided to participants in the discussion group with feedback. Afterwards, the METQ was administered to all participants, which took approximately 10 min to answer. This practice phase was followed by the test phase, and the METQ was used as a distracter between the practice and test phase; the METQ also served as an important part of the study.

Immediate test and delayed tests

On day two, learning was assessed with the immediate test, which featured the items presented in the training sessions. The order of items was altered in this test. The items were displayed in the same manner as in the practice sessions – that is, with a PowerPoint presentation, but without feedback. The students recorded their answers using an answer sheet. At the end of the first retention test (the immediate test), a questionnaire that included some questions about how the discussion groups spent their time during the practice discussions was administered to all discussion groups. The same items and procedure that were used in the immediate retention test were also used in the two delayed tests.

Scoring of the retention tests

To ensure the reliability of the retention test results, the tests were scored by two independent raters and inter-rater reliability was measured with Cohen's kappa statistic (.80) and per cent agreement (92%), suggesting a substantial agreement between the two raters (Watkins & Pacheco, 2000).

Results

The alpha level was set to .05, or .01 when the assumption of equal variance was violated. To estimate effect sizes, we used partial eta squared (η_p^2). Tukey–Kramer *post hoc* tests were used as follow up to the ANOVA analyses.

Retention test performance

One purpose of this study was to examine the learning effects of discussions with or without feedback, as well as of test-enhanced learning on later retention tests. To investigate this, retention test performance was entered into a 3 (Group: discussion group, discussion group with feedback vs. test-enhanced learning) \times 3 (Retention interval: immediate, one week, four weeks) mixed-model ANOVA. Since Mauchly's test showed that the assumption of sphericity was violated, we inspected the multivariate statistics and reported the values of Wilks' Lambda (Λ). The results revealed main effects of group, $F(2, 95) = 14.82$, $p < .0001$, $\eta_p^2 = .24$, and retention interval, $\Lambda = .64$, $F(2, 94) = 25.94$, $p < .0001$, $\eta_p^2 = .35$. These effects were qualified by a statistically significant interaction effect between group and retention interval, $\Lambda = .87$, $F(4, 188) = 3.27$, $p = .013$, $\eta_p^2 = .06$.

To further examine the interaction effect, three separate between-subjects one-way ANOVAs were conducted – one for every test occasion. The groups differed reliably at the immediate test, $F(2, 95) = 20.02$, $MSE = 12.06$, $p < .0001$, $\eta_p^2 = .30$, at the 1-week test, $F(2, 95) = 10.59$, $MSE = 13.89$, $p < .0001$, $\eta_p^2 = .18$, as well as at the 4-week test, $F(2, 95) = 9.53$, $MSE = 12.36$, $p < .0001$, $\eta_p^2 = .17$. Tukey–Kramer *post hoc* comparisons showed that the test-enhanced learning group significantly outperformed the two discussion groups on all three test occasions (Figure 1; immediate test: $p < .0001$ and $p < .0001$,

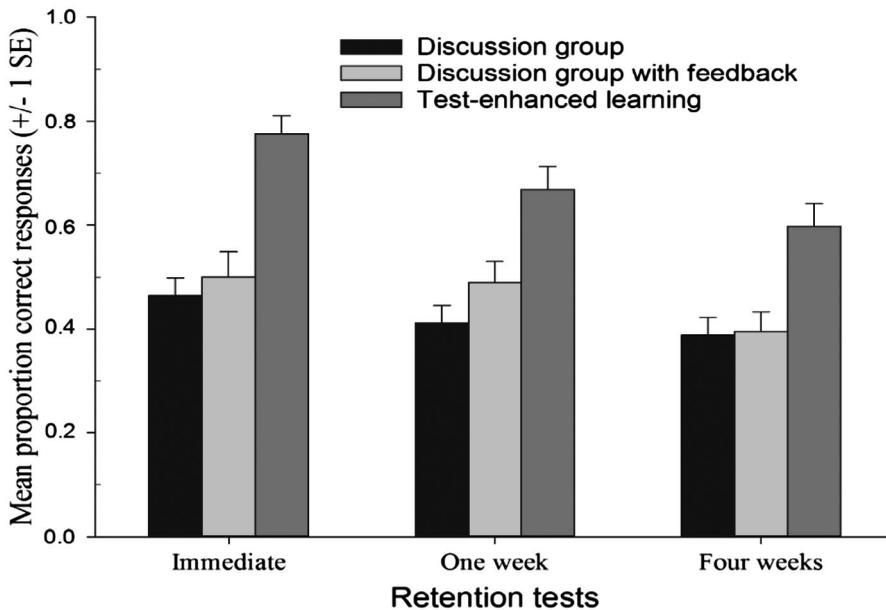


Figure 1. Mean proportion of correct responses for the three groups in the three retention tests. Error bars represent the standard error of the mean.

delayed test one; $p < .0001$ and $p = .008$, delayed test two; $p = .001$ and $p = .001$). As also evident from Figure 1, the two discussion groups did not differ significantly at any retention interval. In summary, the results indicate that feedback does not enhance learning in the discussion groups. Further, it was also found that the test-enhanced learning group consistently outperformed the discussion groups.

To examine whether the type of question (factual or complex) would differentiate the main group effect, group was entered as a fixed factor in a multivariate ANOVA (MANOVA); the composite scores for each type of question were collapsed across all three intervals, which served as the dependent variables. The results of this analysis mimic those of the previous analysis. There was a significant multivariate effect of group, $\Lambda = .68$, $F(4, 190) = 9.19$, $p < .0001$, $\eta_p^2 = .16$. Separate univariate ANOVAs revealed a significant effect of group for both factual, $F(2, 95) = 5.39$, $p < .01$, $\eta_p^2 = .11$, and complex questions, $F(2, 95) = 20.49$, $p < .0001$, $\eta_p^2 = .30$. Tukey–Kramer *post hoc* comparisons showed that the means for the test-enhanced learning group with regard to factual and complex questions ($M = 16.53$, $SD = 4.8$, and $M = 16.12$, $SD = 6.5$, respectively) was significantly higher for both types of questions when compared to the discussion group ($M = 13.11$, $SD = 5.0$, and $M = 7.08$, $SD = 5.1$, respectively) and the discussion group with feedback ($M = 12.90$, $SD = 5.1$, and $M = 9.23$, $SD = 6.3$, respectively). The two discussion groups did, however, not differ significantly.

NFC and learning

To ensure that the groups were comparable with respect to their METQ scores, we first performed a one-way ANOVA with group as the independent variable and the METQ score as the dependent variable. There were no significant differences in METQ scores between the three groups (i.e. discussion with or without feedback, and test-enhanced learning), $F(2, 95) = .46$, $p = .62$, showing that the groups were equal in terms of participants' METQ scores.

To pursue the question of whether the participants' NFC was related to how group discussions affected learning, we divided the participants into two separate groups according to the mean split on their individual METQ scores ($M = 98$, $SD = 15.0$, Median = 99; indicating no problem with extreme

Table 1. Mean (SD) retention test performance at the immediate test, one-week test and four-week test in the three groups as a function of NFC (i.e. METQ) score. The lower part of the table shows the means (SD) merged across the three groups.

Group	Retention tests		
	Immediate	One week	Four weeks
	<i>M</i> (SD)	<i>M</i> (SD)	<i>M</i> (SD)
<i>Test-enhanced learning group</i>			
High NFC	12.53 (3.4)	10.95 (4.1)	9.53 (4.4)
Low NFC	12.23 (3.4)	10.31 (3.9)	9.62 (3.2)
<i>Discussion group</i>			
High NFC	8.53 (3.2)	7.41 (3.6)	7.12 (3.5)
Low NFC	6.39 (2.9)	5.78 (2.6)	5.33 (2.7)
<i>Discussion group with feedback</i>			
High NFC	9.43 (4.4)	9.57 (3.8)	7.36 (4.0)
Low NFC	7.06 (3.2)	6.69 (3.3)	5.63 (2.6)
<i>Total group</i>			
High NFC	10.30 (4.0)	9.36 (4.1)	8.10 (4.0)
Low NFC	8.23 (3.9)	7.34 (3.7)	6.62 (3.3)

values). Thus, the participants in the two discussion groups and the test-enhanced learning group were divided into participants with a high NFC (high METQ scores, >98 , $n = 47$) and those with a low NFC (low METQ scores, <98 , $n = 50$). One participant who received exactly the mean score was excluded from the analysis. The retention test performances were entered into a 3 (group: discussion group, discussion group with feedback, and test-enhanced learning) \times 3 (retention interval: immediate, one week, four weeks) \times 2 (NFC; high, low) mixed-model ANOVA. There was no statistically significant interaction effect for time \times group \times NFC, $\Lambda = .97$, $F(4, 186) = .54$, $p = .703$, nor for group \times NFC, $F(2, 94) = .847$, $p = .432$. There was, however, a main effect of NFC, $F(1, 95) = 5.06$, $p = .027$, $\eta_p^2 = .053$. Those with a high NFC score remembered the information better than those with a low NFC score across all three retention tests. Although the interaction effect of group \times NFC was not significant, the main effect of NFC appears to be driven by a large difference in performance between high and low NFC in the two discussion groups, while the differences in the test group was strikingly similar for all three retention tests (see Table 1).

Discussion

The present study showed that providing feedback to group discussion participants did not aid in the learning and retention of the material at any retention interval, beyond that of the discussion without feedback; moreover, test-enhanced learning with feedback was consistently more effective than group discussions. Most interestingly, the present study also showed that NFC is a personality characteristic that is important to consider when investigating learning. Although the group \times NFC interaction analysis was non-significant, the results indicate that students with high NFC benefit more from participating in group discussions, as opposed to those with low NFC. Conversely, the performance of high and low NFC participants within the test-enhanced learning group was almost identical across all three retention intervals.

Given the previous research, we expected that feedback would enhance learning outcomes after group discussions (Roediger et al., 2001; Vojdanoska et al., 2010). Instead, and somewhat surprisingly, we found that corrective feedback did not have any effect on learning outcomes. A possible explanation for this result could be that the two discussion groups interpreted the questions as being relatively easy, and that the provided feedback did not actually provide any required answer correction beyond

what was already identified (Vojdanoska et al., 2010). Further, the feedback to the discussion groups was provided in written form, and it is possible that the feedback might have been more effective had it been provided orally with a request to also respond orally, thus creating a more active discussion process. This would have likely increased the amount of time participants spent on the questions which, in turn, could have facilitated subsequent performance, as participants would be forced to elaborate on the questions.

We were particularly interested in the relative efficacy of group learning via group discussions when compared with test-enhanced learning. Test-enhanced learning led to better learning than discussion on all three retention intervals, regardless of whether the questions in the retention tests were factual or more complex. Given earlier research, group learning might be expected to be more successful when answering the more complex questions (Kirschner et al., 2009b; van Gog & Sweller, 2015). This result confirms previous research about the beneficial effects of test-enhanced learning, but it provides new and valuable information for both teachers and students on how to most effectively teach or study. The difference might be explained by the amount of time the participants in the discussion groups spent on discussions. The self-rated assessments of the questionnaire indicate that, on average, participants only spent about half the allotted time discussing the questions. It is possible that if participants would have spent more time on the discussions, their performance might have improved. Nevertheless, from a didactical perspective, we believe allocating certain amounts of time to the task at hand mirrors a normal school situation. It is possible that students' experience where 'they know the answers' is based on a feeling of familiarity with the material; conversely, it might be the case that once the participants have perceived that they have already learned the information, they will not allocate any more time to the questions. However, given this study's sample size, it cannot be ruled out that higher power in the analyses would have generated significant differences between group discussions with and without feedback. Hence, there is a risk of a type II error. Another limitation might be that the present study had more female participants than male participants; it is thus possible that an equal number of females and males had generated different results.

There are other plausible explanations that are directly related to comparing test-enhanced learning with group discussions. As was pointed out earlier, there are many studies showing that test-enhanced learning is superior to re-reading the material (i.e. re-study). The effortful retrieval hypothesis has been put forth as a likely explanation for this effect. The basic assumption of the effortful retrieval hypothesis is that more demanding or effortful retrieval during practice strengthens memory, and thus facilitates later recall when compared to studying, which mainly relies on encoding (Bjork, 1994; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013). In the context of the present study, it is possible that participating in group discussions (during which the answers to the questions are always present) bear similarities with re-reading, at least for the participants that are not interested in actively engaging with the material. Test-enhanced learning, on the other hand, requires active engagement with every test question (or at the very least, calls for it).

The transfer-appropriate processing hypothesis is another possible explanation for the beneficial effect of testing over discussion (Morris, Bransford, & Franks, 1977). It states that the more similar the processes and/or contexts between practice sessions and later tests, the greater the final test performance is facilitated (cf. encoding specificity hypothesis; Tulving & Thomson, 1973). In the present study, the practice sessions and subsequent retention tests were identical for the test-enhanced learning group, except for the fact that no feedback was provided during the final retention tests, and that the order of the questions was changed. For the discussion groups, the answers to the questions could be found in the text, which was always at hand (with exactly the same phrasing used for the test-enhanced learning group); the answers could therefore be re-read many times during the group discussion sessions. Although several studies have shown that test-enhanced learning is superior to re-reading, it cannot be ruled out that re-reading interfered with the discussions and therefore negatively affected participants' performance (see Roediger et al., 2001, for a discussion about social influences and false memories).

The analyses of the individual differences in NFC convincingly showed that a person's NFC is important for later performance. As was noted in the introduction, more constrained group discussion appears to be

more beneficial for learning (Morgan et al., 2000). In the present study, both group discussions that used pre-determined questions, as well as the test-enhanced learning condition, have to be regarded as constrained situations. However, the main effect of group on retention performance is still differentiated in terms of NFC. These results suggest that individual differences in motivation – or, as was operationalized by Cacioppo and Petty (1982), to ‘engage in and enjoy thinking’ – is not as important in the context of test-enhanced learning as it might be in other learning contexts, such as group learning. Previous studies have also shown that NFC is associated with intelligence (Hill et al., 2013). Thus, individuals with high intelligence also report high NFC. As intelligence is known to be positively associated with memory performance, this might at least partially explain why those with high NFC outperformed the low NFC group in the retention tests.

Conclusion and further research

The present study shows that written feedback in a group discussion is ineffective, that test-enhanced learning is superior to group discussion, and that NFC is a personality characteristic that affects individual learning outcomes. The results of this study provide unique insights into a problem that is rarely touched upon in school: the use of group discussions under the pretext that it is effective for learning (Morgan et al., 2000). The result that test-enhanced learning was superior to group discussions, regardless of retention interval, whether feedback was provided or not, and NFC level, provides new and valuable information about both learning conditions. Hence, the implication from the present study is that overusing group discussions as a didactical method is not advisable. When a learning strategy (or context) fails, irrespective of whether it is collaboratively or individually organised, the conclusion to draw is that the students or pupils do not need more of that learning strategy; rather, that they need something else (Hattie, 2012, for a review of visible learning). If the target knowledge, as in this case, is to retrieve both simple and more complex factual knowledge (which is a common task in school), it is obvious that test-enhanced learning is a better option and should be utilised to a greater extent in educational settings.

Further research is needed to corroborate the results of the present paper. Thus, this study should be successfully replicated in the same, and different, settings. Furthermore, although the present study indicates that NFC is more important for performance during a group discussion than during test-enhanced learning, the evidence is inconclusive and needs to be examined further. In addition, it is also likely that individuals with higher intelligence also have higher NFC (Hill et al., 2013), and such measures should thus be included in future studies. Moreover, as collaborative learning is argued to facilitate learning, and given that test-enhanced learning in groups has been found to be effective (Cranney, Ahn, McKinnon, Morris, & Watts, 2009), it would be interesting to manipulate the level of interaction between participants within the context of test-enhanced learning.

Acknowledgements

We would like to thank the students and teachers for participating in and assisting with the present research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by Umeå University through the Young Researchers Award and from the Umeå School of Education (Memory and Learning), which was awarded to the third author.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861–876. doi:10.1002/acp.1391

- Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning*, 13, 175–193.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. M. A. P. Shimamura (Ed.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: The MIT Press.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology-Learning Memory and Cognition*, 34, 918–928. doi:10.1037/0278-7393.34.4.918
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116–131.
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119, 197–253.
- Chan, J. C., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology General*, 135, 553–571. doi:10.1037/0096-3445.135.4.553
- Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology*, 51, 291–294.
- Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, 21, 919–940. doi:10.1080/09541440802413505
- Dornic, S., Ekehammar, B., & Laaksonen, T. (1991). Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences*, 12, 313–319.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychology Science in the Public Interest*, 14, 4–58.
- Evans, C. J., Kirby, J. R., & Fabrigar, L. R. (2003). Approaches to learning, need for cognition, and strategic flexibility among university students. *British Journal of Educational Psychology*, 73, 507–528. doi:10.1348/000709903322591217
- Finn, B., & Roediger, H. L. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology*, 39, 1665–1681. doi:10.1037/a0032377
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. London: Routledge.
- Heijne-Penninga, M., Kuks, J. B. M., Hofman, W. H. A., & Cohen-Schotanus, J. (2010). Influences of deep learning, need for cognition and preparation time on open- and closed-book test performance. *Medical Education*, 44, 884–891.
- Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal of Research in Personality*, 47, 22–25. doi:10.1016/j.jrp.2012.11.001
- Jönsson, F. U., Hedner, M., & Olsson, M. J. (2012). The testing effect as a function of explicit testing instructions and judgments of learning. *Experimental Psychology*, 59, 251–257. doi:10.1027/1618-3169/a000150
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19, 528–558. doi:10.1080/09541440601056620
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772–775. doi:10.1126/science.1199327
- Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 704–719. doi:10.1037/0278-7393.33.4.704
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968. doi:10.1126/science.1152408
- Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, 38, 116–124. doi:10.3758/Mc.38.1.116
- Kelley, C. M., & McLaughlin, A. (2012). Individual differences in the benefits of feedback for learning. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54, 26–35. doi:10.1177/0018720811423919
- Kester, L., & Paas, F. (2005). Instructional interventions to enhance collaboration in powerful learning environments. *Computers in Human Behavior*, 21, 689–696. doi:10.1016/j.chb.2004.11.008
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009a). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21, 31–42.
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009b). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25, 306–314.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology*, 35, 989–998.
- Kornell, N., Rabelo, V. C., & Klein, P. J. (2012). Tests enhance learning – Compared to what? *Journal of Applied Research in Memory and Cognition*, 1, 257–259. doi:10.1016/j.jarmac.2012.10.002
- Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2005). Epistemic cooperation scripts in online learning environments: Fostering learning by reducing uncertainty in discourse? *Computers in Human Behavior*, 21, 603–622.
- Maslow, A. H. (1943). Dynamics of personality organization, I & II. *Psychological Review*, 50, 514–539.

- McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*, 200–206. doi:10.3758/Bf03194052
- Meier, E., Vogl, K., & Preckel, F. (2014). Motivational characteristics of students in gifted classes: The pivotal role of need for cognition. *Learning and Individual Differences*, *33*, 39–46. doi:10.1016/j.lindif.2014.04.006
- Morgan, R. L., Whorton, J. E., & Gunsalus, C. (2000). A comparison of short term and long term retention: Lecture combined with discussion versus cooperative learning. *Journal of Instructional Psychology*, *27*, 53–58.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. doi:10.1016/S0022-5371(77)80016-9
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 3–8.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. doi:10.1016/j.tics.2010.09.003
- Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. doi:10.1111/j.1745-6916.2006.00012.x
- Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x
- Roediger, H. L., Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*, 365–371.
- Rydén, G., & Wallroth, P. (2008). *Mentalisering: att leka med verkligheten* [Mentalisation: To play with the reality]. Stockholm: Natur och kultur.
- Slavin, R. E., Hurley, E. A., & Chamberlain, S. (2003). Cooperative learning and achievement: Theory and research. In G. E. Miller & W. M. Reynolds (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 177–198). Hoboken, NJ: Wiley.
- Stenlund, T., Sundström, A., & Jonsson, B. (2014). Effects of repeated testing on short- and long-term memory performance across different test formats. *Educational Psychology*, 1–18. doi:10.1080/01443410.2014.953037
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.
- Van Blankenstein, F. M., Dolmans, D. H. J., Van der Vleuten, C. P. M., & Schmidt, H. G. (2013). Relevant prior knowledge moderates the effect of elaboration during small group discussion on academic achievement. *Instructional Science*, *41*, 729–744.
- van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). The use of textbooks as a tool during collaborative physics learning. *The Journal of Experimental Education*, *69*, 57–76. doi:10.1080/00220970009600649
- van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learning and Instruction*, *12*, 121–138.
- van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage*, *78*, 94–102. doi:10.1016/j.neuroimage.2013.03.071
- van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*, 247–264. doi:10.1007/s10648-015-9310-x
- van Seggelen-Damen, I. C. M. (2013). Reflective personality: Identifying cognitive style and cognitive complexity. *Current Psychology*, *32*, 82–99.
- Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*, 1183–1195. doi:10.1002/acp.1630
- Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education*, *10*, 205–212.