This article was downloaded by:[informa internal users]
On: 10 August 2007
Access Details: [subscription number 755239602]
Publisher: Psychology Press
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format

Online Publication Date: 01 July 2007
To cite this Article: Carroll, Marie, Campbell-Ratcliffe, Jonathan, Murnane, Hannah and Perfect, Timothy (2007) 'Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format', European Journal of Cognitive Psychology, 19:4, 580 - 606
To link to this article: DOI: 10.1080/09541440701326071
URL: http://dx.doi.org/10.1080/09541440701326071

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2007, 19 (4/5), 580–606

Psychology Press
Taylor & Francis Group

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format

Marie Carroll, Jonathan Campbell-Ratcliffe, and Hannah Murnane
Australian National University, Canberra, Australia

Timothy Perfect
University of Plymouth, Plymouth, UK

Retrieval practice leads to the impaired recall of related but unpractised items, an effect termed retrieval-induced forgetting (RIF). Two experiments showed that RIF occurred with "real world" educational material, and isolated several boundary conditions for the phenomenon. Experiment 1 showed that integration of material available to experts but not to novices in a domain of knowledge, protected against RIF, which disappeared after a 24 hours. Experiment 2 examined the impact on RIF of the degree of coherence or integration of the text material itself and the type of test format administered. Text coherence did not influence RIF, which occurred for the short answer and essay tests, but not for the multiple choice test. In both experiments, those participants who demonstrated RIF were able to monitor accurately the likelihood of recall impairment, suggesting that RIF may not be an unconscious process. Results are discussed in relation to exam preparation strategies.

When competition is introduced between items that are to be learned, participants can be unconsciously induced to forget those items that compete for retrieval (Anderson, Bjork, & Bjork, 1994). For example, if participants are given two categories of items to learn, and then perform sustained retrieval practice on half of the items from one of the categories, their later recall of unpractised competitors from the practised category is found to be impaired, relative to their recall of items from the wholly unpractised category. This form of unconscious forgetting is termed RIF, and has been found to be quite robust across a large range of studies.

Correspondence should be addressed to Marie Carroll, QESS, Australian National University, Canberra, ACT 0200, Australia. E-mail: marie.carroll@anu.edu.au

© 2007 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business
http://www.psypress.com/ecp DOI: 10.1080/09541440701326071

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Investigations of this effect have touched on several important practical implications, ranging from how students study for exams (and may unwittingly forget studied information), to the effect of interviews (and thus retrieval practice) on eyewitness testimony, to the effects of forgetting in social psychology. For instance, Macrae and MacLeod (1999) demonstrated that RIF could be elicited under mock examination conditions, while MacLeod (2002) and Shaw, Bjork, and Handal (1995) found RIF was the cause of impaired recall in a modified eyewitness paradigm.

Many studies (including Anderson, Bjork, & Bjork, 2000; Anderson & Spellman, 1995; Bäuml & Kuhbandner, 2003; Dunn & Spellman, 2003; Johnson & Anderson, 2004; Perfect, Moulin, Conway,& Perry, 2002; Saunders & MacLeod, 2002; Shivde & Anderson, 2001; and see Levy & Anderson, 2002, for a review) have argued that RIF can be best accounted for by inhibitory processes. Specifically, it is argued that while repeated retrieval practice clearly leads to the strengthening of practised items, those items that are semantically related, and will later compete for activation during retrieval, are inhibited to better improve the recall of those items that are most appropriate for completing the task.

The cue used to elicit inhibited items may be varied without affecting the inhibition effect. As shown by Anderson and Spellman (1995), the cue–target relationship alone is not the issue—forgetting is highly dependent on the competing items, and specifically their contextual relationship to the practised items.

The present studies are concerned to investigate the occurrence of RIF in formal classroom learning, especially in relation to students studying for exams. There are several important boundary conditions for RIF that have particular implications in classroom learning. The first boundary condition is the influence of integration of learned material on RIF, the second is the duration of the RIF effect, and the third is the effect of type of test format that is most susceptible to RIF. Together, these boundary conditions may make RIF more or less likely to be observed in classroom learning, and are the focus of the investigation in the following studies.

INTEGRATION OF LEARNED MATERIAL

A factor that appears to moderate the amount of RIF experienced is the degree to which participants are able to integrate the practised items into a cohesive whole. Generally speaking, integration refers to the existence of interconnections between items sharing a common retrieval cue. These interconnections can be long standing (i.e., existing preexperimentally), or novel (i.e., generated during the course of an experiment). Studies of fact recognition initially suggested that integration may protect against RIF. For

Downloaded By: [informa internal users] At: 16:35 10 August 2007

example, Radvansky and Zacks (1991) instructed participants to learn a series of facts (e.g., "the palm is in the lobby", "the ashtray is in the lobby"). Generally, the more facts the participants learned, the slower their recall of specific facts (termed the "fan effect"). But if participants are able to integrate these facts (in the above example, the facts are grouped by their common location), little interference is found. Several studies indicate that integration can also reduce RIF, and, interestingly, it appears that little needs to be done to encourage participants to integrate items while studying and performing retrieval practice. Anderson and McCulloch (1999) found, in a simple modification of the Anderson et al. (1994) paradigm, that an initial instruction to "integrate information" was enough to significantly reduce the amount of RIF observed during the final test phase. When Anderson and McCulloch later divided their participants into high and low integrators, depending upon their responses in postexperimental questionnaires, they found that participants who reported that they had spontaneously attempted to integrate the material, even when not specifically required to, also demonstrated significantly less RIF. In a further demonstration, Anderson and Bell (2001) found that integration through the formation of single vivid mental images protected against the forgetting of propositional materials (e.g., "the ant crawled on the table", "the ant crawled on the rock"). However, it is important to note that while both of these studies found that participants who integrated the material demonstrated a reduction in the amount of RIF, it rarely completely eliminated it. Rather, the amount of integration reported by participants tended to be associated with a proportionate reduction in the amount of RIF observed during the final testing phase.

The findings of Dunn and Spellman (2003) also favour the role of integration as a boundary condition for RIF. In this study, while practising some stereotypic traits (e.g., Artist–creative, or Asian-American–intelligent) inhibited the recall of unpractised traits relating to that stereotype, this effect was strongly modified by the degree to which the participant believed in the stereotype. Dunn and Spellman argued that, as a stereotype provides an integrative framework for understanding the world, participants who strongly believed in the validity of the stereotype were able to cohesively integrate the stereotyped traits. This integration reduced the amount of interference between competing traits, thus reducing the degree of inhibition and the amount of forgetting.

As Anderson and McCulloch (1999) point out, the effect of RIF is somewhat paradoxical when we consider expertise. Assuming that learning more information about a topic increases competition, then as we acquire more expertise within a particular area we should become *less* adept at retrieving items from within this domain. If the inhibitory processes behind RIF represent an adaptive attempt to reduce competition, then surely an

Downloaded By: [informa internal users] At: 16:35 10 August 2007

expert's large domain of knowledge should be particularly prone to inhibition and thus forgetting. However in reality, it seems that experts generally have *better* access to information from within their respective domains. In an attempt to understand this paradox, Smith, Adams, and Schorr (1978) proposed that experts may integrate knowledge from within their domain, and this integration may be what protects against forgetting. Experts do not accumulate new facts in isolation, but rather integrate them within a rich body of background knowledge that supports an understanding of the many interrelations between items. Anderson and McCulloch (1999) argue that this integrative framework encourages the understanding of interrelations between items and the development of novel retrieval strategies, which may protect experts from the paradoxical nature of RIF. To date, however, there have been few studies of the operation of RIF using complex information in experts' domains of knowledge. One such study recently conducted (Barnier, Hung, & Conway, 2004) did find RIF for emotional and unemotional autobiographic memories, knowledge, which could be construed as coherent and organised in the same way that expertise is. However, further investigation is necessary to determine whether high levels of integration in experts are capable of reducing competition between items, and thus protect them from RIF in classroom settings.

One study that has examined the educational implications of RIF is that of Macrae and MacLeod (1999), who subjected participants to a mock exam scenario, in which they were presented with facts relating to two fictional tropical islands (Tok and Bilu). They then performed retrieval practice on half of the facts relating to one of the islands. Consistent with RIF, recall of unpractised facts about the practised island was impaired relative to the recall of facts relating to the unpractised baseline island. However, in this study the "fact" to be learned was explicit, with equal weight given to each (i.e., "Bilu's only major export is copper"). There was also no rich fabric of background knowledge on which to draw to support learning. In contrast, in real-world learning contexts, it is up to the student to read through a large body of information, to draw inferences based on background and textual knowledge, and to select ideas likely to be critical (and likely to be tested later) from those which are less important. It is unclear whether RIF would occur under these circumstances.

DURATION OF THE RIF EFFECT

The literature consistently agrees that inhibition in the retrieval-practice paradigm will be evident from 5 (Macrae & MacLeod, 1999) to 20 min (Anderson et al., 1994, among others) after retrieval practice is performed.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Several studies have indicated that RIF will recover over time. MacLeod and Macrae (2001, Study 1) tested participants either 5 min or 24 hours after performing retrieval practice, and while RIF was observed immediately following retrieval practice, it was not evident after 24 hours. This result indicates that somewhere in this time period, the inhibition of competing items was reduced to zero. In their Study 2, the authors placed a 24 hour delay between either retrieval practice and the final test, or between study and retrieval practice. Interestingly, this manipulation produced the expected forgetting only among participants in the second condition, indicating that a delay of 24 hours only eliminated RIF when it occurred *after* retrieval practice, and not after encoding, although the magnitude of forgetting was not as great as when all phases followed closely.

METAMEMORY MONITORING

A further unexplored issue is whether individuals are aware of the memory impairments that RIF generates. Dunn and Spellman (2003) argue that lateral inhibition is responsible for the RIF effect, and is automatically activated when attention is drawn to one category or another. Unlike hierarchical inhibition, which involves the conscious and intentional inhibition of items via top-down methods of processing (thought to be responsible for directed forgetting), it is an automatic process. Even so, this does not necessarily imply that it is always outside consciousness. Surprisingly, to our knowledge, no research has investigated self-monitoring of RIF—specifically whether individuals are aware that performing retrieval practice on some items may impair the later recall of related but unpractised items. There are educational implications here. If students are aware that once-learned but unpractised information is likely to be inhibited, is there anything they can do to compensate for this? The very retrieval practice that students are repeatedly encouraged to perform before tests—and which has been shown to improve test performance (e.g., Roediger & Karpicke, 2005)—may make their memory for some items worse. It is normally impracticable to rehearse all of the information one needs to know about a topic in preparing for an exam; usually only a subset is subject to retrieval practice. Yet there may be an unwarranted assumption on the learner's part that recently learned unrehearsed information related to recently retrieved information is as likely as any other once-learned information to be remembered. Although performing compensatory activities may not be an option, it is at least useful to educators to know whether learners are aware of the inhibitory effects on unpractised items. Research has demonstrated that individuals revise learning strategies in light of the content they feel has and has not been consolidated into memory (Nelson & Leonesio, 1988;

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Pelegrina, Bajo, & Justicia, 2000); yet, it is unclear whether memory inhibition may influence these judgements. Judgements of the success and progress of an individual's own learning are functions of an individual's metacognitive monitoring (Nelson & Narens, 1994; Schwartz & Perfect, 2002). One such judgement is the Judgement of Learning (JOL), typically an assessment made after the learning phase and prior to final testing, which aims to assess people's degree of confidence that they will correctly recall the information that they have learned at a later time (Nelson & Narens, 1994). If individuals in the retrieval-practice paradigm indicate that they are less confident in their ability to recall the unpractised items (Rp−) from the practised category (Rp+) than to recall unpractised items from the baseline category (NRp items), this may indicate that RIF is not a wholly unconscious phenomenon.

Our primary aim in these studies was to investigate whether integration can protect against forgetting. In the first study we examined the hypothesis put forward by Anderson and McCulloch (1999) that experts may be protected against forgetting information from their domain of expertise through an ability to integrate new information into a cohesive framework. In the second study we investigate integration by varying the type of material presented. In both cases, the material used was real-world classroom text to be learned for an upcoming test.

EXPERIMENT 1

In Experiment 1, the "experts" were students who had studied psychology formally at university for 4 or more years; the novices were those in the first year of university psychology study. The two case studies were taken from university textbooks describing schizophrenia and autism respectively (but with no reference to these labels). It was thought that experts would readily integrate the symptom set into one of the two disorders, while novices would be less able to label and, therefore, integrate the symptom set into a preexisting schema for the disorder, though they may recognise a subset of symptoms as typical of one or the other disorder.

We examined the occurrence of RIF for such material when testing took place immediately and when it was delayed by 24 hours. While MacLeod and Macrae (2001) found that RIF did not occur when retrieval practice and final testing were separated by 24 hours, the duration of inhibition may depend greatly upon the type of material that is being inhibited (Anderson, 2003). Category-exemplar pairs may be inhibited for a different length of time than real-world material, which may in turn be different from that found for the simple, artificial "facts" of the MacLeod and Macrae study. To date no study has examined the effect of RIF on complex real-world

Downloaded By: [informa internal users] At: 16:35 10 August 2007

material, and thus its persistence is unknown. For practical purposes, a finding that persistence of forgetting occurs after 24 hours is of concern to students and teachers, since many students study only in the 24-hour period before an exam.

Additionally, Experiment 1 included JOL ratings for the material that had been learned. These were elicited following the retrieval-practice phase for both rehearsed and unrehearsed items in order to determine whether learners were able to monitor RIF, should it occur.

Method

Participants and design

There were 32 participants in each of two categories: novices and experts. The novices, half of whom were first year Psychology students, had a mean of 0.38 years of psychology experience. For the expert participants the mean number of years of psychology experience was 4.55 ($SD = 0.76$), and all were Psychology majors.

This experiment involved a $2 \times 2 \times 2$ mixed factors design, where type of item (two levels, Rp− vs. NRp[1]) was manipulated within subjects, while the degree of expertise (two levels: novice vs. expert) and the time period between retrieval practice and final testing (two levels: immediate vs. 24 hours) were between-subject factors. The dependent variables were: (1) the number of questions answered correctly and (2) the mean JOL rating.

Materials

Two case studies with 40 related questions were constructed, with content drawn largely from a third-year abnormal psychology textbook. One case study was of a schizophrenia sufferer named David, and the other was of an autism sufferer named Peter. The texts were rewritten as vignettes apparently describing particular cases, so as to ensure that all the information was entirely novel, and could not be predicted without having first read the papers. For each case study, 20 specific and distinct questions were devised (see Appendix A for examples). An effort was made to include a range of easy and difficult questions, but the emphasis was on including more difficult questions to eliminate ceiling effects amongst experts. During the final testing phase, each participant was presented with only 30 of the total 40 questions. These 30 comprised 10 practised questions (Rp+), 10 unpractised questions from the practised case study (Rp−), and 10 baseline questions from the unpractised case study (NRp).

[1] An explanation of NRp and Rp– follows in the Procedure section.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Participants were provided with an answer booklet, with two variations. The first pages consisted of 30 rating scales intended for the assessment of JOLs. These scales ranged from 0 to 100 in increments of 20, with 0 indicating no confidence of recalling the answer at final testing, and 100 indicating full confidence. In the immediate condition a distractor task followed, comprising three pages containing each letter of the alphabet written sequentially, and a category heading at the top of the page. This section was absent from answer booklets used in the 24-hour condition. On the following pages, numbered spaces allowed participants to write their answers for all 30 questions in the final testing phase.

Procedure

Participants were randomly assigned to one of the testing conditions. The following procedure follows the original Anderson et al. (1994) retrieval-practice paradigm closely, but with the addition of a JOL phase.

Initial study phase. Participants were instructed to read the two study papers carefully. It was made clear that they would be tested on their comprehension of these papers later with a series of short answer questions, and were instructed to "pick out any points which they considered salient". Presentation of the papers was counterbalanced—half of the participants read the David/schizophrenia paper first, while half read the Peter/autism paper first.

Retrieval-practice phase. Participants were then read a series of 10 questions, which they were told would appear on the final test. This phase was intended to mimic the practice, which is common prior to an exam, where examinees quiz one another on material that they may be tested on. Participants were instructed to respond verbally with the answer if they knew it, and to guess even if they were unsure. If they were incorrect or did not know the answer, they were required to repeat the correct answer, which was provided by the experimenter. This repetition was critical as it was intended to ensure retrieval of the material, as opposed to simple re-presentation, which may not have produced the desired RIF effect (Bäuml, 2002). Participants practised half of the questions from their assigned set, and each question was repeated three times. Participants were randomly assigned to practise either the odd or even questions from one of the sets for counterbalancing purposes. These practised questions became the Rp+ items for that participant, the unpractised questions from the same case study became their Rp− items, and 10 odd or even items from the unpractised case study became their NRp items.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Judgements of learning. Next, participants were shown each question, without the answer, that would appear in their final test, and for each they were instructed to indicate their level of confidence that they would correctly recall the answer during the final testing phase. It was emphasised to participants that they should avoid thinking of the answers themselves, to avoid a retrieval practise opportunity for the unpractised items. To reduce the likelihood of covert retrieval, participants were instructed to respond as quickly as they could, and were reminded of this every time their response was delayed.

Distractor phase. Those participants assigned to the "immediate" condition engaged in a series of distractor tasks, lasting 15 min, which involved generating 26 words beginning with each letter of the alphabet, for three categories. Participants in the "24-hour" condition were instructed to return at the same time on the following day for the final phase of the experiment.

Final test phase. Participants were provided with a question booklet containing their 30 counterbalanced questions, and were instructed to answer as many as they could in their answer booklet. Participants were told that they would not lose any marks for incorrect answers, and thus to try to answer as many questions as they could.

Results

Initial screening. A significance level of .05 was used for all statistical tests. It was possible that certain combinations of questions were easier or harder than others, which could affect the amount of inhibition demonstrated by participants with these combinations of questions. To determine if this was the case, two one-way ANOVAs were conducted, the first assessing whether the choice of practised case study affected the amount of inhibition, and the second assessing whether practising odd or even questions affected the amount of inhibition. Neither was found to be significant, $F(1, 62) = 3.081$, $p = .084$; and $F(1, 62) = 0.391$, $p = .534$, respectively.

Proportion correct. Retrieval practice did facilitate later recall, as can be seen from the means for Rp+ items in Table 1. Because the large, predicted practice effect swamps other effects, the analyses that follow did not include Rp+ items.

Several researchers have argued that RIF is measured as the amount of inhibition each participant demonstrates (see, for example, Perfect et al., 2002). Inhibition is calculated for each participant by subtracting Rp− scores from NRp scores separately for each condition. Table 1 shows the

Downloaded By: [informa internal users] At: 16:35 10 August 2007

TABLE 1
Mean proportion of questions answered correctly in each experimental condition in Experiment 1

Condition	*Rp+*	*Rp−*	*NRp*	*Inhibition (RP−)−(NRp)*
Novice/immediate	.894	.338	.506	−.169
Novice/24 hour	.763	.356	.394	−.038
Expert/immediate	.956	.525	.575	−.050
Expert/24 hour	.869	.519	.538	−.019

mean scores on amount of inhibition for each of the four conditions. The 2 (novice vs. expert) × 2 (immediate vs. 24 hour) × 2 (Rp− vs. NRp) ANOVA showed that, not surprisingly, experts performed better (mean = 0.53, standard error = 0.28) than novices (mean = 0.39, standard error = 0.29), $F(1, 60) = 11.98$, $p < .001$, $MSE = 5.28$, and that NRp items (mean = 0.50, standard error = 0.22) were answered more correctly than Rp− (mean = 0.43, standard error = 0.22), $F(1, 60) = 13.18$, $p < .001$, $MSe = 15.12$, confirming an overall RIF effect. The main effect of delay did not reach significance (mean immediate = 0.48; mean 24 hour = 0.45). The interaction between the two factors—novice/expert and type of question—did not reach significance, $F = 3.29$, $p > .05$; however, the levels of inhibition presented in Table 1 suggest that novices experienced greater RIF than experts in the immediate condition. Experts, then, are protected somewhat from RIF.

There was also a significant interaction between delay and type of question, $F(1, 60) = 4.61$, $p < .05$, $MSe = 1.47$; after 24 hours the RIF effect was much less pronounced than in the immediate condition. The three-way interaction was not significant.A separate analysis was carried out using inhibition (the mean difference, for each subject, between Rp− and NRp) as the dependent variable. As it was hypothesised that inhibition would be greater in the novice/immediate condition, a simple planned comparison was executed. The resulting $F(3, 60) = 9.303$, $p < .05$ and $R^2 = .134$ indicated that the amount of inhibition in the novice/immediate condition did indeed differ significantly from the other three conditions, supporting the hypothesis that RIF could be produced for this naturalistic material.

As it was specifically predicted that novice participants in the immediate condition would demonstrate greater inhibition than novices in the 24-hour condition, a final planned comparison was executed. This revealed that there was indeed a significant difference between these conditions, with $F(3, 60) = 6.007$, $p < .05$ and $R^2 = .091$, indicating significantly more inhibition amongst participants in the immediate condition.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

TABLE 2
Mean proportion of JOL ratings by item type for each condition, Experiment 1

Condition	*JOL+*	*JOL−*	*JOLN*	*Inhibition (JOL−)−(JOLN)*
Novice/immediate	.94	.52	.56	−3.50
Novice/24 hour	.79	.53	.60	−6.25
Expert/immediate	.98	.70	.65	+4.50
Expert/24 hour	.98	.68	.69	−1.25

Magnitude of judgements of learning. A similar 2 (novice/expert) × 2 (delay) × 2 (Rp−/NRp) ANOVA was carried out on the judgements of learning; the means are shown in Table 2. Again, items that had been practised (Rp+) received much greater JOL ratings; for this reason Rp+ items were not included in the analysis. There were no significant main effects or interactions for JOL ratings. The Novice/expert × Type of item interaction failed to reach significance, $F = 3.53$, $p = .06$.

Normally, gamma correlations between the JOL rating and the response accuracy are calculated for each subject. However, with only 10 questions per condition (Rp+, Rp−, and NRp) there were many empty cells, and the gamma correlations were not able to be calculated.

It appears, from the JOL magnitude results that participants were not able to predict the significant inhibition that occurred; that is, they did not know that Rp− items would be answered less accurately than NRp items. Nevertheless, though the interaction was not significant, the means suggest that the novices were more likely to predict the inhibition effect than the experts. Perhaps, then, those subjects in general who are more susceptible to RIF are more likely to predict a poorer performance on Rp− than on NRp items.

To assess this possibility, participants were divided into two groups—those who demonstrated inhibition (those whose Rp− scores were lower than their NRp scores), and those who did not. Of the total 64 participants, 36 (or 56%) demonstrated some degree of inhibition. The mean proportion for JOLs for each question type across these two groups is presented in Table 3.

The amount of predicted inhibition (lower ratings to Rp− than to NRp items) was compared to determine whether those participants who demonstrated RIF also predicted this inhibition as shown by the proportions of JOLs. This was indeed found to be the case, $F(1, 62) = 11.22$, $p < .001$, $Mse = 85.15$, indicating that significantly more inhibition was predicted by those participants who demonstrated retrieval-induced forgetting.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

TABLE 3
Mean proportion of JOLs for each question type for those participants who did and did not demonstrate RIF, Experiment 1

	n	*JOL+*	*JOL−*	*JOLN*	*Inhibition (JOL−)−(JOLN)*
*P*s with inhibition	36	.94	.57	.63	−6.4
*P*s without inhibition	28	.91	.66	.62	+4.57

Discussion

This study aimed to improve our overall understanding of the mechanisms involved in RIF, and to provide fresh directions for the development of learning strategies in real educational settings based upon this information. Our first goal was to determine whether RIF would occur for "real-world" material in a mock exam scenario, as previous research has focused on forgetting in laboratory settings using fairly artificial materials. Our second was to assess the notion of Anderson and McCulloch (1999): that integration may afford experts some protection from the detrimental effects of inhibition. Our third was to determine whether inhibition of this material would decay after 24 hours, as previous research has demonstrated (MacLeod & Macrae, 2001; Saunders & MacLeod, 2002). And finally, we sought to determine whether individuals have some level of awareness of the inhibition incurred by retrieval practice in this paradigm, or whether forgetting is entirely unconscious.

Each of these hypotheses was supported by the results. As was expected, the only condition that demonstrated significant levels of forgetting was the novice/immediate condition—participants in the other conditions were protected from forgetting by the effects of integration, or a sufficient time interval, which allowed the inhibition to decay. These results strongly suggest that RIF will occur for this real world material, but with two notable exceptions: (1) Domain knowledge affords protection from forgetting when the material can be integrated into a preexisting framework, and (2) a gap of 24 hours between retrieval practice and final testing is sufficient to reduce the forgetting of Rp− items to negligible levels. More intriguingly, these results also indicate that those individuals who experience inhibition may have some awareness of the detrimental effects of retrieval practice. This implies that the inhibition produced by retrieval practice may not be entirely unconscious. We return to this point in Experiment 2.

When viewed as a whole, the significantly smaller amounts of inhibition observed in the expert/immediate condition compared to the novice/immediate condition implies an advantage for those with some background in the material. There is some uncertainty about whether this advantage is

Downloaded By: [informa internal users] At: 16:35 10 August 2007

directly due to integration of the material. Anderson and McCulloch (1999) observe that experts are better able to recall individual items from within a comparatively larger body of information, and while they argue that integration is the mechanism responsible, "integration" is a construct that is difficult to tap into directly, leaving the possibility that other models may also account for the reduced levels of forgetting observed among experts. For instance, it could be argued that experts have simply had more practice at recalling items from within their domain, and thus their improved recall of individual items may be a result of more frequent practice at recalling this type of material. Indeed, the experts have simply had more practice in study and test-taking in any domain; it is possible that it is the experience at university, rather than the domain expertise, that confers the protection. Future studies might aim to exercise more control over "expertise" using non-Psychology, senior students as experts.

Previous studies (MacLeod & Macrae, 2001; Saunders & MacLeod, 2002) have found that RIF is almost eliminated after 24 hours. This study provided support for these findings, with no significant amounts of inhibition evident after a delay of 24 hours between retrieval practice and final testing. However, it is possible that different types of material may produce different levels of inhibition that may persist for differing lengths of time. Presenting material in prose format may encourage integration over material presented as disconnected sentences; it is possible that the rate at which inhibition recovers may not be the same for these different types of material.

Interestingly, while our results have indicated that RIF is eliminated by a 24-hour break between retrieval practice and testing, facilitation of practised items when compared to nonpractised items was still apparent after this period. This observation raises an interesting question: Why is it necessary to inhibit competing items to facilitate the recall of practised items in the short term, yet after a 24-hour delay the facilitation of practised items remains but without the inhibition of unpractised items? It is possible that different mechanisms are involved in these differing time periods. Thus inhibition may represent a relatively short-term strategy for improving the recall of items against their stronger competitors, but when given a significant temporal delay, the relative strengthening of the practised items alone is sufficient to improve their recall whilst avoiding the false recall of competing items.

It is clear that material of this type is not immune to the detrimental effects of retrieval practice. While RIF is an issue neither for long-term learning nor for material being learned within a domain of expertise, it is still the case that in some classroom learning situations, new material may be learned, practised, and finally quizzed within a single learning session. Such a teaching practice would be highly inadvisable based upon these findings, as it would almost certainly result in significant levels of RIF.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

EXPERIMENT 2

Experiment 1 found that RIF is a potentially serious limitation on test performance, at least for novices in a knowledge domain, and that its effect is most serious immediately after learning. Nevertheless, learners do seem to be aware of its existence, and can therefore take steps to overcome it.

In Experiment 2 we extended our examination of its effects on real-world educational materials to investigate another type of integration, which may afford protection against RIF. This integration is that provided by the form of presentation of the material itself; coherent text may, by virtue of its continuity, provide more protection from RIF than a series of unrelated statements. Reading sentences in an ordered or a disordered fashion influences the degree to which the material is integrated in memory (Carroll & Korukina, 1999; Shaddock & Carroll, 1997). Experiment 1 showed that coherently presented text is subject to RIF, but did not investigate whether the same text presented as a series of disordered statements is *even more* subject to RIF. There are educational applications in understanding this. For instance, when studying for an exam, students may vary the degree to which they integrate learned material. For some students, a learning strategy that entails summarising material into bullet points that capture separate elements of the information may suffice; for other students, material may need to be integrated into an overarching conceptual framework (commonly known as "deep" learning). As stated above, integrating material during the learning phase has been demonstrated as a factor that can protect Rp− items from inhibition (Anderson & Bell, 2001; Anderson & McCulloch, 1999). With instructions that encouraged participants to either rehearse each item with previously studied items from the same category or to simply study the relationship between each category-exemplar word pair (such as fruit–orange), Anderson and McCulloch (1999) found that encouraging the integration of exemplars reduced RIF. In the classroom, the analogous situation is the presentation of material to be learned in an ordered (textually coherent) or a disordered fashion.

Experiment 2 aims to investigate both integration and another variable, type of test format, to determine the relevance of RIF to real-world learning, particularly in exam situations. Initial investigations failed to report RIF with recognition tests (Anderson & Bjork, 1994; Butler, Williams, Zacks, & Maki, 2001). Butler et al. (2001) observed RIF with a category-cued recall test, but not with a fragment-completion test, a word-fragment-cued recall test, a category-plus-fragment-cued recall test, and a category-plus-stem-cued recall test. It may be that item-specific cues, such as word fragments, release the inhibition impairing that item's recall (Butler et al., 2001). These conclusions have since proven to be inconclusive (Anderson, 2003), with other studies demonstrating RIF on tests of item recognition (Hicks

Downloaded By: [informa internal users] At: 16:35 10 August 2007

& Starns, 2004; Verde, 2004). Clearly, there is a need for a systematic investigation of the susceptibility to RIF of variations of test format typically found within educational settings, such as multiple choice, short answer, and essay-style tests. An additional departure from the standard RIF paradigm is that in real-world study settings, and in our study, the questions, multiple choice distractors, and required targets may not exactly repeat the wording of the material learned in the study phase. If RIF were observed with such materials, it would further strengthen the notion that memory inhibition is cue independent.

Experiment 1 found that JOL predictions are influenced by the memory inhibition of unpractised material, suggesting a degree of awareness about the results of inhibition. Judgements of learning were lower for Rp− items than for NRp items for those participants who demonstrated RIF effects. A review of a number of studies and theoretical discussions provides inconclusive evidence as to whether recall predictions can distinguish between those items that have and have not been inhibited in memory (Benjamin, Bjork, & Schwartz, 1998; Carroll & Shanahan, 1997). The addition of a judgement of learning before testing, but after retrieval practice, is included in Experiment 2 to determine whether the apparent awareness of inhibition is a reliable finding.

Thus, Experiment 2 aims to (1) examine the integration-protection hypothesis by manipulating type of material—integrated or not integrated; (2) determine whether RIF is apparent with real-world material that is tested in a variety of classroom test formats; and (3) replicate the finding of Experiment 1, that, where RIF is apparent, individuals can predict the impairment in Rp− items before the testing takes place.

Method

Participants

Forty-five undergraduate Psychology students with a mean age of 21.96 years ($SD = 6.28$, range of 18–47 years) were tested individually.

Design

The design was a 3 (retrieval-practice condition of learned information: Rp+, Rp−, and NRp) × 2 (text integration: ordered and disordered) × 3 (test format: multiple choice, short answer, and essay) mixed model. Retrieval practice and text integration were manipulated within participants. Test format was manipulated between participants and provided three experimental groups, with repeated measures on text integration creating a total of six experimental conditions, with 15 participants in each condition.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

The inclusion of a within-participants test of the integrated text variable in the present study follows the suggestions of Carroll and Nelson (1993) who showed that such designs are more sensitive in detecting the effect of an independent variable on metacognitive judgements.

The dependent measures were (1) percentage correct recall and (2) magnitude of Judgement of Learning (JOL) ratings. These measures were calculated across each of the six conditions.

Materials

Four different texts were adapted from introductory psychology textbooks; care was taken to ensure that each did not overlap considerably with general knowledge. The selected topics were Electroconvulsive Shock Therapy (Text A), Mental Disorders (Text B), Self-Actualisation (Text C), and Defining Intelligence (Text D). Each heading was centred at the top of the page as the text "category". Each text was approximately 500 words and an A4 page in length. For the disordered condition, the sentences from the original texts were arranged in alphabetical order, and those beginning with a pronoun were expanded to state the replaced noun.

Retrieval-practice materials. Two sets of retrieval-practice materials, each containing five statements, one or two sentences long, were created for each of the four texts, with a "category" heading at the top of the page. Some of the statements for practice comprised information from across two sentences. Retrieval practice consisted of the participant completing the missing phrase in each statement in response to its first letter, and thus constituted a cued-recall test. For example, for the original sentence: *Together, the first four needs are referred to as deficiency needs*, practice consisted of supplying the missing word for the frame: *Together, the first four needs are referred to as d_____ needs.* Participants were required to successfully retrieve the missing phrase for each statement three times, in its entirety.

Test materials. A multiple choice, short answer, and an essay-style test was created for each of the four texts, with each test having its "category" heading at the top of the first page. These tests aimed to replicate those given in educational contexts, and differed in the amount of self-initiated retrieval required. The multiple choice test presented participants with one correct answer and three incorrect but plausible alternative responses, most of which were from the teaching instruction manual accompanying the textbook. Appendix B shows parallel versions of the same questions in different test formats.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

The short answer questions were the same as the multiple choice questions; each required a single unique response, and no half marks were given for incomplete answers. Pilot testing revealed that all questions were of similar difficulty level.

The essay-style test asked participants to recall as much information as they could remember from the text they had just read. They could list the information in bullet-point form if preferred. In this case, the task was less like writing an essay, and more similar to preparing a plan for an essay-style exam response. Nevertheless, for convenience, it is referred to as the essay test condition.

JOL rating materials. JOL ratings were given for each of the 10 questions per text, after the rehearsal practice had occurred and before the test. Judgements were made on a 6-point rating scale, where 1 = "*I feel I have not learned the information and will not recall it at a test in five minutes*" and 6 = "*I feel I have learned the information well and will certainly recall it correctly in five minutes*". Importantly, JOLs were made to the question only, without the answer.

Tasks. Four distractor tasks were used, two relating to verbal tasks, and two to numerical tasks.

Procedure

Individual testing sessions lasted from 60 to 70 min. Counterbalancing the order of presentation of the texts and test formats was achieved using a stratified random sample based upon the 24 combinations of text materials, A, B, C, and D. In the learning phase, participants read the first text for 5 min knowing that their recall (test unspecified) of the information would be assessed later in the study. To replicate real study conditions, they were allowed to make notes during the study, but they could not access these at test. After 5 min, they read a second text for the same purpose.

Retrieval practice commenced immediately after completion of the learning phase. Five statements from one of the two texts they had just read (the Rp+ items) were read aloud. Each had one missing word cued by the first letter, to be completed by participants. If they were unable to complete a word, immediate feedback was given on their performance, with the intention of having participants correctly complete each statement three times. All statements had to be answered correctly three times nonconsecutively. No participant required more than five attempts to reach this criterion. Following retrieval practice, participants performed a distractor task for 3 min and then provided JOL ratings for the 10 questions associated with each text read during the learning phase. These questions comprised the

Downloaded By: [informa internal users] At: 16:35 10 August 2007

five Rp+ and five Rp− items from the rehearsed text, and the ten NRp items from the unrehearsed text. Only the questions, not the answers, were presented in the JOL ratings phase, in order to prevent a further retrieval practice opportunity.

A second 3-min task followed JOL ratings, after which participants were given separate 5-min tests on the two texts they had just read. There was no knowledge of type of test format until this point, to ensure that all participants learned the material in the same way.

Upon completion of the two tests, the procedure was repeated using two different texts, with these texts being different in format (ordered or disordered) to the first two.

Results

Dependent measures consisted of (1) proportions correct and (2) JOL rating magnitude for Rp+, Rp−, and NrP information. To test for the main results of interest, the inhibition effect, mixed-model repeated measures analyses of variance were carried out with the factors being 2 (text order: ordered vs. disordered) × 3 (test format: multiple choice, short answer, essay) × 2 (item type: Rp− vs. NRp).

Participants correctly completed 83.13% ($SD = 9.01$) of the missing words for each statement during retrieval practice across conditions. This figure is comparable to that of previous researchers (Anderson et al., 1994; Anderson, Green, & McCulloch, 2000; Anderson & McCulloch, 1999; Dunn & Spellman, 2003; Macrae & McLeod, 1999) and indicates the success rate of retrieval practice in the present study.

Proportion correct. The mean recall performance of participants according to item type, test format, and text order conditions are presented in Table 4. There was a significant effect of text order, $F(1, 42) = 4.93$, $p < .05$, $Mse = 521$, with ordered text being better recalled (mean = 0.43, $SEm = 2.53$) than disordered text (mean = 0.36, $SEm = 2.70$). There was also a significant inhibition effect, $F(1, 42) = 10.9$, $p < .01$, $Mse = 309$, with the mean performance on Rp− items (0.35, $SEm = 2.7$) poorer than that of NRp items (mean = 0.44, $SEm = 1.97$). The test format effect was also significant, $F(2, 42) = 27.94$, $p < .001$, $Mse = 717$, with the mean for multiple choice format (0.59) being significantly greater than that of short answer (0.36) and essay (0.23) formats. The only significant interaction was that between inhibition and test format, $F(2, 42) = 4.56$, $p < .01$, $Mse = 309$. Pairwise comparisons revealed no difference between Rp− and NrP item recall in the multiple choice condition, indicating no RIF effect with this type of test format. In the short answer condition, Rp− recall ($M = 30.67$,

Downloaded By: [informa internal users] At: 16:35 10 August 2007

TABLE 4
Mean proportions of correct and standard deviations by test format, and order of text, for Rp+, Rp−, and NRp items, Experiment 2

	Ordered text			*Disordered text*		
Type of test	*Rp+*	*Rp−*	*NRp*	*Rp+*	*Rp−*	*NRp*
Multiple choice	.77	.62	.63	.72	.60	.54
	(.82)	(.58)	(.50)	(.80)	(.66)	(.45)
Short answer	.69	.40	.50	.64	.21	.35
	(.82)	(.58)	(.50)	(.80)	(.66)	(.45)
Essay	.52	.14	.33	.44	.17	.28
	(.82)	(.58))	(.50)	(.80)	(.66)	(.45)

$SE = 4.52$) was significantly lower ($p < .05$) than NrP recall ($M = 43$, $SE =$ 3), and in the essay condition, Rp− recall ($M = 15.33$, $SE = 2.91$) was also significantly lower ($p < .001$) than NrP recall ($M = 31.33$, $SE = 3.33$). Together, these analyses indicate the operation of RIF with the short answer and essay tests, but not in the multiple choice condition.

Judgements of learning. Mean JOL ratings according to item type, test format, and text integration conditions are presented in Table 5. A 2 (text order) × 2 (item type: Rp− vs. NRp) × 3 (test format) repeated measures ANOVA on JOL magnitude revealed only one significant main effect: that of text order, $F(1, 42) = 14.23$, $p < .001$, $Mse = 0.81$. Ordered text was judged to be more likely to be remembered (mean = 3.6, $SEm = 0.12$) than disordered text (mean = 3.11, $SEm = 0.12$). No interactions were significant.

As in Experiment 1, then, participants were not able to predict the significant inhibition that occurred; that is, they did not know that Rp− items would be answered less accurately than NRp items.

TABLE 5
Mean proportion of JOL ratings and standard errors by test format, and order of text, for Rp+, Rp−, and NRp items, Experiment 2

	Ordered text			*Disordered text*		
Type of test	*Rp+*	*Rp−*	*NRp*	*Rp+*	*Rp−*	*NRp*
Multiple choice	.378	.357	.376	.369	.329	.33
	(.29)	(.26)	(.25)	(.33)	(.26)	(.25)
Short answer	.432	.357	.402	.384	.299	.304
	(.29)	(.26)	(.25)	(.33)	(.26)	(.25)
Essay	.474	.346	.334	.393	.298	.301
	(.29)	(.26)	(.25)	(.33)	(.26)	(.25)

Downloaded By: [informa internal users] At: 16:35 10 August 2007

As in Experiment 1, a further analysis was conducted separately for those participants who showed RIF and those who did not show RIF. As Table 6 shows, the mean JOL scores for those participants who showed RIF mirrored the proportion correct scores.

It would be desirable to be able to compare Rp− and NRp conditions for each test format for participants who experienced exactly the same combinations of test questions. However, due to the small numbers, it was not possible to make the comparisons having regard to the counterbalancing conditions. Thus, it is possible that the split into high and low levels of inhibition was due to combinations of items that were particularly difficult or easy. Thus, any conclusions about the awareness of RIF must be regarded as tentative. When individual paired *t*-tests were performed between Rp− and NRp items for those participants who showed RIF, there were no significant effects in any of the three test format conditions, due, no doubt, to the small sample sizes.

These results are promising but not firm (for the reasons given above) evidence that people are able to monitor the products of the inhibition effect. Further specifically designed studies with larger sample sizes are needed.

TABLE 6
Mean proportion of JOL predictions according to the presence or absence of RIF across test format conditions, Experiment 2

		Rp+		*Rp−*		*NRp*	
	N	*Mean proportion correct*	*Standard error*	*Mean proportion correct*	*Standard error*	*Mean proportion correct*	*Standard error*
RIF participants							
Multiple choice	8	.388	.042	.343	.043	.399	.032
Short answer	14	.400	.029	.308	.036	.322	.027
Essay	18	.450	.031	.304	.024	.353	.025
Non-RIF participants							
Multiple choice	22	.369	.027	.343	.019	.341	.019
Short answer	16	.415	.025	.346	.024	.380	.027
Essay	12	.410	.042	.350	.026	.265	.023

Results have been collapsed across text order, creating a total sample of 90 observations from 45 participants.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Discussion

Experiment 2 had three aims: to compare the operation of memory inhibition in classroom-like learning situations under conditions where the material itself was presented in an integrated or an isolated format; to investigate whether RIF occurs with different test formats; and to determine whether individuals can monitor the inhibited information. This study demonstrated that the degree of integration of the text did not provide the degree of protection against RIF that was observed in the experts in Experiment 1, suggesting that, at least for newly learned material, it is irrelevant whether the information presented is integrated or not.. The results also showed that RIF occurred in both short answer and essay tests, but not in the multiple choice test; and that individuals who experience RIF may perhaps be able to monitor the effect of that inhibition.

Previously, Anderson and colleagues (Anderson & Bell, 2001; Anderson & McCulloch, 1999) have demonstrated that encouraging the integration of learned category-exemplar word pairs reduces RIF. The idea here is that integration increases the interconnections between items and helps to transfer the benefits of retrieval practice for Rp+ items to the unrehearsed Rp− items. In the Anderson studies, participants were instructed to form interconnections based upon similarities and differences between word pairs, a manipulation not explicitly required in our study, although it would be expected this would occur automatically in integrated text.

Some studies that provide an alternative explanation for RIF—the strategy disruption hypothesis—would suggest that ordered text should be more susceptible to RIF than disordered text. According to the strategy disruption hypothesis (Basden & Basden, 1995; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003) RIF may occur because retrieval practice disrupts the original organisation of the learned information in memory. While Rp+ items benefit from retrieval practice, Rp− items become more difficult to remember. NrP items are easier to recall than the Rp− items because of a lack of practice and disruption (MacLeod et al., 2003). Recalling information that has been learned in a disordered and isolated fashion may not be susceptible to the disruption resulting from retrieval practice. Hence, RIF may be reduced for a disordered text. Indeed in our study, the Rp− versus NRp difference *was* greater (though not significantly so) in both the ordered short answer condition and in the ordered essay condition than in the disordered versions. Despite the plausibility of the strategy disruption hypothesis, an explicit test of this hypothesis is yet to be undertaken in reference to RIF.

With inconsistencies in the literature surrounding the operation of RIF with recall versus recognition testing (Anderson & Bjork, 1994; Butler et al., 2001; Veiling & van Knippenberg, 2004; Verde, 2004), no clear prediction

Downloaded By: [informa internal users] At: 16:35 10 August 2007

was made concerning RIF's interaction with test format. Our results indicated an absence of RIF for the multiple choice test in both order conditions. In the short answer and essay tests, however, RIF was clearly apparent. Essay and short answer tests differ from multiple choice tests in the degree of self-initiated recall required for correct answers: In multiple choice tests, selecting the correct response is more a process of recognition (Oosterhof, 2003) than retrieval. A failure to observe RIF with recognition, but not a recall, test is consistent with the results of Butler et al. (2001) and Koustaal, Schacter, Johnson, and Galluccio (1999); the latter argued that the retrieval cues are enough to overcome the suppression of the nonreviewed Rp− items, with such cues reexposing the individual to the initial learning episode. Such an explanation suggests that inhibition is a flexible process that can be overcome with item-specific cues. Other studies *have* found RIF with tests of item recognition when the dependent measure was reaction time latencies (Veiling & van Knippenberg, 2004). Observing RIF with reaction time latencies (Veiling & van Knippenberg, 2004) and not with correct response rates in recognition tests (Butler et al., 2001; Koustaal et al., 1999) suggests that Rp− items are inhibited before the completion of a recognition test. The inclusion of reaction time latencies and confidence ratings (for accuracy of lures and targets) within the retrieval-practice paradigm may clarify this possibility. In addition, these measures may also indicate the extent to which RIF was masked by guessing in the present study, given that multiple choice tests are somewhat susceptible to guessing (Oosterhof, 2003).

It may also be the case that our results were confounded by the choice of distractors used in Experiment 2. These distractors were not items that had been exposed in the text. An alternative way of doing the multiple choice test would be to include distractors that were studied earlier, as is often the case in real educational settings. Perhaps our Rp− items in multiple choice were indeed inhibited, but not to the point where they were indistinguishable from completely new items. The familiarity of the target answers among the new distractors in our study may have been responsible for the apparent absence of RIF. Some caution should be exercised, then, in concluding that multiple choice tests are not affected by RIF, especially in light of the studies mentioned above (Hicks & Starns, 2004; Verde, 2004) that RIF has been observed in recognition memory tests.

The other line of enquiry in Experiment 2 concerned metamemorial monitoring of the inhibition effect. The failure of participants to accurately detect Rp− inhibition was at odds with the finding of Experiment 1, where participants accurately monitored the Rp− versus NRp difference. However, when participant data were analysed according to whether or not an individual showed RIF, the findings were generally consistent with Experiment 1. The impracticability of comparing subjects who received exactly the

Downloaded By: [informa internal users] At: 16:35 10 August 2007

same combinations of items in the RIF and non-RIF groups tempers any conclusions we are able to draw, but if RIF monitoring is possible, then there are implications for classroom learning, such as that students studying for an exam can assess their poor performance on some items with some reliability. In considering exactly what the participants are monitoring, we are indebted to a reviewer (B. Levy, personal communication) who suggests that what people are tapping into is the results or products of the inhibition, rather than the inhibition itself. Awareness of the process of inhibition is not necessary: They simply register that the Rp− item, for whatever reason, seems harder to recall than other items.

Our results were consistent with previous work on monitoring of integrated text. Previous research has demonstrated that JOLs are sensitive to both the amount of learning (Shaddock & Carroll, 1997) and text coherence and context (Carroll & Korukina, 1999; Mazzoni & Nelson, 1995). In both instances, individuals predict greater recall for overlearned items and for material learned in an ordered sequence. In the present Experiment 2, the integrated text did indeed receive significantly higher JOL ratings than the nonintegrated text.

GENERAL DISCUSSION

We have established that RIF is a robust phenomenon that occurs with real-world classroom material standard in formal learning situations. It is apparent that material of this type is not immune to the detrimental effects of retrieval practice. We have also established a boundary condition on the protection that integrated material affords against RIF. In Experiment 1, integration of material that was provided by expertise in the domain did indeed provide this protection. The background knowledge of the experts effectively meant that the newly learned material was not inhibited in the same way as it was for novices. This accords with Anderson and McCulloch's (1999) observation that experts are better able to recall individual items from within a comparatively larger body of information. In Experiment 2, a different method of inducing "integration" of the material was manipulated, one that relied on presentation format, rather than individual expertise, as a means of inducing integration. This involved comparing expository prose format, which itself may encourage spontaneous integration of the material, with a series of disparate and apparently unconnected facts. Experiment 2 found that presenting information to students in a holistic way, which might better encourage integration, did not protect them from RIF.

It appears that individuals can indeed monitor the effect RIF will have on their later test performance, at least in short answer and essay tests.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Experiment 1 also supports previous studies that have found that inhibitory effects reduce dramatically over the course of 24 hours. In terms of real-world applications for this data, this study suggests that RIF is not an issue for long-term learning, being certainly a transitory effect. Furthermore, in common tertiary-level learning scenarios where material is presented and learned in a holistic and integrated manner, RIF is not likely to be an issue. However, in some educational contexts, it is conceivable that new and unconnected material may be learned, practised, and finally quizzed within a single learning session. Such a teaching practice would be highly inadvisable based upon these findings, as it would almost certainly result in significant levels of RIF.

Despite pressures from teachers and parents to effectively practise and review material in the lead-up to an exam, an individual's exam performance may be impaired by this process. In some circumstances an individual's exam performance may be better off without excessive overlearning (see also Nelson & Leonesio, 1988) of some items at the expense of others. As suggested by Macrae and MacLeod (1999), retrieval practice may indeed be detrimental to exam performance, particularly when the test format is short answer or essay. However, where retrieval practice of a subset of items is encouraged as a study technique, being aware of the likely outcomes of retrieval practice may allow students to compensate in some way prior to a test.

REFERENCES

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General*, *130*(3), 544–570.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). RIF: Evidence for a recall-specific mechanism. *Psychonomic Bulletin and Review*, *7*(3), 522–530.

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 265–326). San Diego, CA: Academic Press.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087.

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1141–1159.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on RIF. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 608–629.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *120*(1), 68–100.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Barnier, A. J., Hung, L., & Conway, M. A. (2004). RIF of emotional and unemotional autobiographical memories. *Cognition and Emotion*, *18*(4), 457–477.

Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1656–1669.

Bäuml, K.-H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science*, *13*, 356–360.

Bäuml, K.-H., & Kuhbandner, C. (2003). RIF and part-list cuing in associatively structured lists. *Memory and Cognition*, *31*(8), 1188–1197.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on RIF. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1314–1319.

Carroll, M., & Korukina, S. (1999). The effect of text coherence and modality on metamemory judgements. *Memory*, *7*(3), 309–322.

Carroll, M., & Nelson, T. O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subjects than in between-subjects designs. *American Journal of Psychology*, *106*, 227–235.

Carroll, M., & Shanahan, C. (1997). The effect of context and metamemory judgements on automatic processes in memory. *Acta Psychologica*, *97*, 219–234.

Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, *39*, 420–433.

Hicks, J. L., & Starns, J. J. (2004). RIF occurs in tests of item recognition. *Psychonomic Bulletin and Review*, *11*(1), 125–130.

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science*, *15*(7), 448–153.

Koustaal, W., Schacter, D. L., Johnson, M. K., & Galluccio, L. (1999). Facilitation and impairment of event memory produced by photograph review. *Memory and Cognition*, *27*(3), 478–493.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Science*, *6*(7), 299–305.

MacLeod, M. D. (2002). RIF in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology*, *16*, 135–149.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. *Psychology of Learning and Motivation*, *43*, 163–214.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of RIF. *Psychological Science*, *12*(2), 148–152.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*(3), 463–473.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 1263–1274.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labour-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalf & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Oosterhof, A. (2003). *Developing and using classroom assessments* (3rd edn). New York: Pearson Education.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

Pelegrina, S., Bajo, M. T., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory*, *8*(6), 377–392.

Perfect, T. J., Moulin, C. J., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of RIF with implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1111–1119.

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*, 940–953.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of RIF in misinformation effects. *Journal of Experimental Psychology: Applied*, *8*(2), 127–142.

Schwartz, B. L., & Perfect, T. J. (2002). Introduction: Toward an applied metacognition. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 1–14). Cambridge, UK: Cambridge University Press.

Shaddock, A., & Carroll, M. (1997). Influences on metamemory judgements. *Australian Journal of Psychology*, *49*(1), 21–27.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). RIF in an eyewitness-memory paradigm. *Psychonomic Bulletin and Review*, *2*(2), 249–253.

Shivde, G., & Anderson, M. C. (2001). The role in inhibition in meaning selection: Insights from retrieval inducted forgetting. In D. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 175–190). Washington, DC: American Psychological Association.

Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology*, *10*, 438–464.

Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 315–318.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory and Cognition*, *32*(8), 1265–1272.

APPENDIX A

1. What did teachers first notice was odd about David's behaviour? Talked to his dead uncle.
2. Why was David unable to finish high school or obtain a job? He was agitated and verbally aggressive.
3. Which of David's delusions is a particular form of delusion known as Capgras syndrome? His belief that his aunt had been replaced by a double.
4. How could David's conversational speech be described? Highly disorganised.
5. David's tendency to jump from topic to topic and talk illogically is known by psychologists as? Associative splitting.
6. How did David demonstrate inappropriate affect? He laughed when thinking about the death of his uncle.
7. What was David's most common form of auditory hallucination? Hearing his dead uncle.

Downloaded By: [informa internal users] At: 16:35 10 August 2007

APPENDIX B

Examples of alternative forms of the same questions in Experiment 2.

Text: Electroconvulsive Shock Therapy

Extract from the text:

> "When ECT is applied in the traditional manner bilaterally (with the current running across both of the brain's hemispheres), the patient typically loses memory for events that occurred a day or two before the treatment."

Multiple choice

What type of memory loss may be associated with bilateral ECT?

A. Events that occurred the day of treatment
B. Pictorial memories
C. Events occurring a day or two before treatment
D. Verbal memories

Short answer

What type of memory loss may be associated with bilateral ECT?

Free recall

Please recall as many points that you can remember and learned about when reading about Electroconvulsive Shock Therapy. You may list your points in bullet-point form on this sheet of paper (both sides if required).

Note: Retrieval practice for this item was of the following form:

> When ECT is applied in the traditional manner bilaterally (with the current running across both of the brain's hemispheres), the patient typically loses memory for e_____ that occurred a d___ o__ t___ before the treatment.