Using electronic flashcards to promote learning in medical students: Retesting versus restudying

6 AUTHORS, INCLUDING:

Ralf Schmidmaier
Ludwig-Maximilians-University of Munich
55 PUBLICATIONS **836** CITATIONS

SEE PROFILE

Miriam Schiller
Ludwig-Maximilians-University of Munich
4 PUBLICATIONS **24** CITATIONS

SEE PROFILE

Inga Hege
Geisel School of Medicine at Dartmouth
50 PUBLICATIONS **216** CITATIONS

SEE PROFILE

Martin R Fischer
Ludwig-Maximilians-University of Munich
261 PUBLICATIONS **1,490** CITATIONS

SEE PROFILE

Using electronic flashcards to promote learning in medical students: retesting versus restudying

Ralf Schmidmaier,[1] Rene Ebersbach,[1] Miriam Schiller,[1] Inga Hege,[1] Matthias Holzer[1] & Martin R Fischer[1,2]

CONTEXT The superiority of retesting over restudying in terms of knowledge retention and skills acquisition has been proven in both laboratory and classroom settings, as well as in doctors' practice. However, it is still unclear how important retesting strategies are to the learning of relevant factual knowledge in undergraduate medical education.

METHODS Eighty students in Years 3–5 of medical school in Munich participated in a prospective, randomised, double-blinded, controlled study in which they were exposed to 30 electronic flashcards designed to help them memorise key factual knowledge in the domain of clinical nephrology. The flashcards were presented in four successive learning cycles, each consisting of a study period and a subsequent test period. Half of all participants were randomised to repetitive studying (restudy group) and half were randomised to repetitive testing (retest group) of successfully memorised flashcards. Knowledge retention was assessed after 1 week and 6 months.

Additionally, personal data, self-reflection on the efficacy of the learning strategies and judgements of learning were obtained by questionnaires.

RESULTS Repetitive testing promoted better recall than repetitive studying after 1 week (p < 0.001). However, after 6 months general recall was poor and no difference between the restudy and retest groups was observed. Time on task and number of trials, in addition to sex, age, performance and psycho-social background, did not vary between the groups. Self-predictions of student performance did not correlate with actual performance.

CONCLUSIONS In the context of using electronic flashcards, repetitive testing is a more potent learning strategy than repetitive studying for short-term but not long-term knowledge retention in clinical medical students. Although students use testing as a learning strategy, they seem to be unaware of its superiority in supporting short-term knowledge retention.

[1]Medical Education Unit, Medizinische Klinik Innenstadt, University Hospital, Ludwig-Maximilian University (LMU), Munich, Germany
[2]School of Medicine, Faculty of Health, Witten/Herdecke University, Witten, Germany

Correspondence: Dr Ralf Schmidmaier, Schwerpunkt Medizindidaktik, Medizinische Klinik Innenstadt, Klinikum der Universität, Ludwig-Maximilians-Universität München, Ziemssenstrasse 1, Munich 80336, Germany.
Tel: 00 49 89 5160 2111; Fax: 00 49 89 5160 4410;
E-mail: ralf.schmidmaier@med.lmu.de

INTRODUCTION

Given the large and growing body of relevant biomedical knowledge, the development of effective and efficient learning strategies is a key and classical challenge for medical students. The use of flashcards represents an ubiquitous study strategy,[1,2] including among medical students,[3,4] which is becoming increasingly popular.[2,4,5] Flashcards are mainly used for the acquisition of the factual knowledge that represents a central component of medical education, the assimilation of which is a prerequisite for clinical reasoning. The efficacy of clinical reasoning methods used by experts depends on the expert's possession of a highly elaborated knowledge base[6,7] and reasoning experts struggle in attempts to solve problems in other domains in which they lack sufficient knowledge.[8,9]

Flashcards are usually used for studying and subsequent self-testing. Recent research in cognitive psychology has shown that testing *per se* affects learning. Although it has long been recognised that 'assessment drives learning'[10,11] indirectly by increasing motivation to study, by generating a greater amount of preparation activity, by stimulating improvement in study strategies and by facilitating the individual's discovery of his or her weaknesses and encouraging subsequent restudy,[12,13] recent work has begun to show that tests also have direct effects on knowledge retention.[12–16] The retrieval of information that occurs when one is tested appears to strengthen the memory of this information,[17] thereby increasing the likelihood that the information will be successfully retrieved on subsequent tests.[18]

A variety of studies have shown that taking a test is superior to restudying for an equivalent amount of time, thus underpinning the efficiency of this approach.[13,17,19,20] Testing seems to be especially effective in promoting long-term knowledge retention.[21,22] However, students often appear to be unaware of these benefits when they are asked to rate the effectiveness of various learning strategies.[17,23] These findings have predominantly been established in laboratory settings using subjects in which participants have no prior knowledge in order to control confounding variables. A series of studies conducted at college level demonstrated that such laboratory findings could be replicated in the classroom,[14] thus proposing testing as a robust and powerful approach to promote learning. In medical education, testing in combination with training has also proven to be more effective than training alone with reference to the learning of basic life support skills.[24] Another recently published study[25] conducted in a medical residency programme demonstrated that repetitive testing spaced over time can also be used as a powerful learning tool in a professional clinical setting to improve long-term knowledge retention. This educational context is not congruent with that of medical students, who have only limited prior knowledge of a specific topic and usually do not make direct connections between factual knowledge and specific aspects of patient care.

However, the concept of repetitive testing as a superior strategy to improve knowledge retention has high potential and needs to be investigated for medical education in ecologically valid settings. Medical students seem to be a highly select group with high cognitive capacity and often have a particular psycho-social background.[26,27] Therefore, it is still unknown and questionable whether the results of laboratory or classroom experiments with psychology students or undergraduate college students can be applied in undergraduate medical education.

In the present study, we developed a set of domain-specific electronic flashcards and evaluated their feasibility and usability in a retesting versus a restudying context in clinical medical students. Furthermore, we examined whether retesting is more efficient than restudying in terms of its ability to foster short- and long-term knowledge retention. Finally, we assessed the ability of participants to monitor the metacognitive effects of retesting versus restudying experiences.

METHODS

Learning materials

Thirty flashcards were developed. Each included a title (cue) and corresponding items (targets). The 30 flashcards contained 98 items in total, resulting in a range of two to five items (mean: 3.3 items) per flashcard. The medical content referred to factual information on clinical nephrology and more precisely to acute renal failure and chronic renal insufficiency. The contents of the flashcards were linked across the cards. Appendix S1 (online) shows the content. An example flashcard is presented in Fig. 1(a). The contents represented the information on which clinical reasoning and decision making are based in the management of patients with acute and chronic renal failure. The electronic flashcards were created and displayed using the CASUS™ online
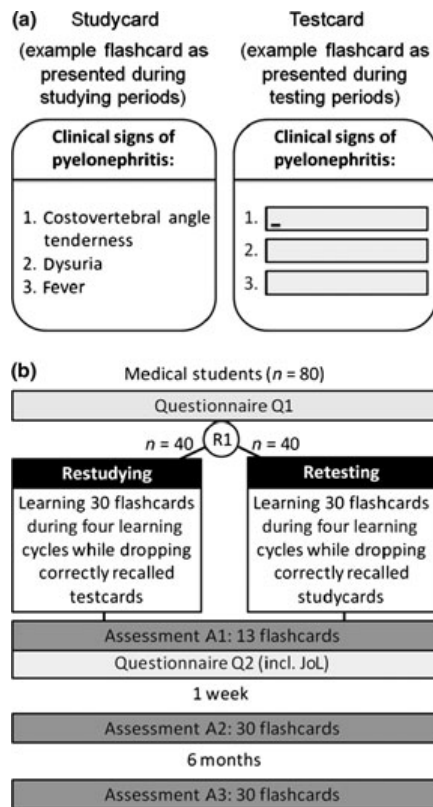
Figure 1 (a) Example flashcards as presented on a CASUS™ screen during study and test periods. (b) Study design. See Appendix 1 for details defining the restudying and retesting conditions.

learning environment, which has proven effective in various studies.[28,29]

Participants

Eighty students in Years 3–5 of medical school at Ludwig-Maximilian University (LMU) Munich volunteered to participate in the study. Participants were stratified by gender and study semester prior to randomisation. Students were randomised on a 1 : 1 basis by a computer program to two groups (n = 40 per group). The study was carried out in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants. The ethical committee of LMU Munich approved the study. Study participants received compensation for expenses.

Study design

All subjects filled out an anonymised questionnaire (Q1) containing items on their psycho-social background, gender, age, motivation, mood, frequently used learning strategies and prior knowledge.

A standardised introduction was given to all participants (Appendix S2). Participants were asked to memorise the items on the flashcards as well as possible. All participants were made aware that learning success would be assessed. Participants were then randomised 1:1 to either the restudy group or the retest group (detail defining the 'restudying' and 'retesting' conditions is illustrated in Appendix 1).

Participants then logged into the electronic learning (e-learning) environment and embarked on the learning phase designated to the group to which they had been randomised (retesting or restudying). Investigators and participants were blinded to the group allocation. The students' knowledge retention was compared between the groups at 1 week and 6 months according to the amount of flashcard content they correctly recalled (Fig. 1b).

Learning phase: repetitive testing versus repetitive studying

The e-learning environment used by each group included the same 30 nephrology flashcards, but the way in which correctly recalled flashcards were repeated differed between the two groups. Learning comprised four cycles in each group. Each learning cycle consisted of a study period followed by a distractor task and a test period. During study periods, students viewed each flashcard including its title (cue) and learning objectives (targets) on one screen of the CASUS™ system (Fig. 1a). Participants studied in a self-directed manner; no time limitation was imposed, but time on task was recorded. During testing trials, students viewed the cue and a cursor (Fig. 1a) and were asked to type in the correct targets in the form of open-ended text (for two to five items per flashcard; a maximum Levenshtein distance of 2 was allowed to accommodate typographical errors). The number of expected targets was indicated. Again, no time limit was imposed, but time on task and answers given were recorded electronically. No feedback was given. In the first learning cycle all 30 flashcards were displayed for studying. After a subsequent distractor task, all 30 cues were given to test for the corresponding targets. If all targets to a cue were typed in correctly by the participant, he or she was considered to have successfully acquired the knowledge content of that flashcard. Methods of fostering knowledge retention differed between the two randomised groups. In the restudy group, correctly recalled flashcards were studied repetitively and the corresponding test cards were dropped from subsequent learning cycles. Accordingly, in the retest group, correctly recalled

flashcards were tested repetitively and the corresponding study cards were dropped from subsequent learning cycles.

When they had completed the four learning cycles, participants were assessed using 13 test cards (Assessment 1 [A1]).

At the end of day 1, students were asked to predict how many of the 30 flashcards they would be able to recall after 1 week (representing their judgement of learning [JoL]). They then completed a questionnaire (Q2) asking for their assessment of the efficacy of repetitive studying versus that of repetitive testing. Participants were then dismissed.

Assessment of knowledge retention in the two groups

Participants returned for a short-term knowledge retention test 1 week later. On this test (Assessment 2 [A2]), subjects were shown each of the 30 flashcard titles (cues) and were told to type in the correct answers (targets). Six months later, 76 of the initial 80 participants (95%) were available to take part in an assessment of long-term knowledge retention (A3). Again the cues of all 30 flashcards were provided and participants were asked to type in the 98 target items.

Statistics

We calculated that a group size of 40 participants would allow for an absolute difference of > 10% with $\alpha = 0.01$ and power = 0.80. P-values were calculated by t-test in cases of normal distribution and otherwise by the Wilcoxon rank sum test. P-values of < 0.05 were considered to indicate statistical significance. Cohen's d as effect size was calculated by means and standard deviations (SDs).

RESULTS

Study population and the e-learning tool

Participants (57% females) were currently studying in their eighth semester. Their mean age was 25.2 years (SD = 5.5). Their mean score on the first national board examination after Year 2 of medical school was 2.55 (SD = 0.79) (1 = very good, 5 = insufficient) for achieving a mean of 75.16% (SD = 8.37%) correct items. Fifty-four students had already completed assessments in internal medicine; the average grade in the internal medicine examination was 1.97 (SD = 0.85) (1 = very good, 5 = insufficient). Overall, 34% of participants had

considered specialising in internal medicine and 4% suffered from a kidney disease or had a relative with a nephrological problem. Characteristics including college grades and education and profession of parents were similar between the two study groups, with no significant differences. In their responses to Q2, 76 of the 80 participants reported that they had 'learned much today', corresponding to a median response of 2.0 (interquartile range [IQR] 1.0) on a 5-point Likert scale (1 = strongly agree, 5 = strongly disagree). Using the same 5-point Likert scale, the majority of respondents indicated that they 'would use the computer program for self-directed study' (median = 2.0, IQR 1.75).

Acquisition of knowledge using electronic flashcards: learning curves and time on task

Although the primary goal of the study was to assess the retention of descriptive knowledge on nephrology, the influence on learning of repetitive testing versus repetitive studying was also analysed more closely. As correctly answered flashcards were eliminated in the subsequent testing period, the speed and success of learning could be assessed according to learning condition (retesting versus restudying). Figure 2 shows that the learning curves are congruent with a slight, non-significant trend towards faster acquisition of knowledge in the restudy group. However, in A1, administered after the distractor task at the end of the first experimental day, both groups were able to correctly recall the complete contents of 70% of the flashcards (retest group: 69.98%; restudy group: 70.48%). When scores for correct items on flashcards that were not completely recalled are added, the retest and restudy groups achieved scores of 86.97% and 88.56%, respectively.
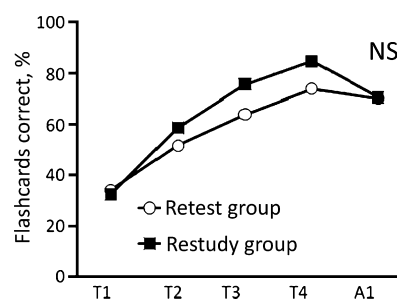
Figure 2 Analysis of knowledge acquisition. Cumulative performance during the learning phase was assessed by the percentage of correctly answered test cards during the four subsequent test periods (T1–T4) and after a distractor task (A1). Mean percentages of correctly answered test cards are shown for the retest and restudy groups, respectively. NS = not significant

Table 1 Average number of trials within each studying (S) or testing (T) period, average time on task per period, and total number of trials in the learning phase in each condition. Each screen of the e-learning tool represents one trial

Period	Learning cycles								Total
	1		2		3		4		
	S1	T1	S2	T2	S3	T3	S4	T4	
Restudy group									
Trials	30	30	30	20	30	12	30	7	189
Time, min	16	24	11	21	7	19	4	18	120
Retest group									
Trials	30	30	20	30	15	30	11	30	196
Time, min	17	27	13	19	9	17	6	15	123

By contrast with the protocols used in previous studies, time on task was not limited. Nevertheless, both groups used the same amount of time and trials (Table 1). There was a slight but not significant trend towards the use of more trials (relative difference 2.7%) and increased time on task (relative difference 2.6%) in the retest group, which probably reflects the somewhat flatter learning curve shown in Fig. 2.

Knowledge retention at 1 week and 6 months

The study hypothesis declared that repetitive testing would prove superior to repetitive studying with respect to the retention of knowledge. Figure 3 demonstrates that subjects in the retest group performed significantly ($p < 0.001$) better than those in the restudy group on A2 at 1 week by recalling 40.00% of flashcards correctly, compared with the 27.75% of flashcards recalled correctly by the restudy group; these scores represent the correct recall of 67.20% and 57.30% of all items, respectively. The calculated Cohen's *d* effect size is 0.76, indicating a medium-to-large difference.

After 6 months, 76 of the initial 80 medical students undertook a third assessment (A3), which was identical to A2 given at 1 week. Cues for all 30 flashcards were given and participants were required to type in the target items. The advantage of the retest group over the restudy group was lost at this long-term

follow-up (Fig. 3). Both groups correctly recalled only about 11% of the flashcards (retest group: 10.33%; restudy group: 11.42%; difference not significant), corresponding to 34.28% and 38.29% of all items, respectively. The time on task increased between assessments by about 10 seconds, but did not significantly differ between the two experiment groups: the time taken per flashcard was 28.47 seconds (retest group) and 34.47 seconds (restudy group) at A1, 42.89 seconds (retest group) and 45.90 seconds (restudy group) at A2, and 54.74 seconds (retest group) and 52.98 seconds (restudy group) at A3.

The study was powered to investigate differences between the percentage of complete flashcards and percentage of flashcard items recalled correctly. In A1, results showed a substantial absolute difference of 18% between the proportion of flashcards recalled correctly (70%) and the proportion of flashcard items recalled correctly (88%), indicating the relative inferiority by 20% of flashcard success compared with item success (18%/88%). This increased to an absolute difference of 28% and relative inferiority of 45.6% after 1 week (A2) and an absolute difference of 25.4% and relative inferiority of 70.0% at 6 months (A3). In summary, over time our participants lost most of the factual knowledge they had acquired so that the percentages of complete flashcards and flashcard items they were able to recall correctly dropped from approximately 70.2% and 87.8%, respectively, at A1, to 33.9% and 62.3% and finally to 10.9% and 36.3%, at A2 and A3,
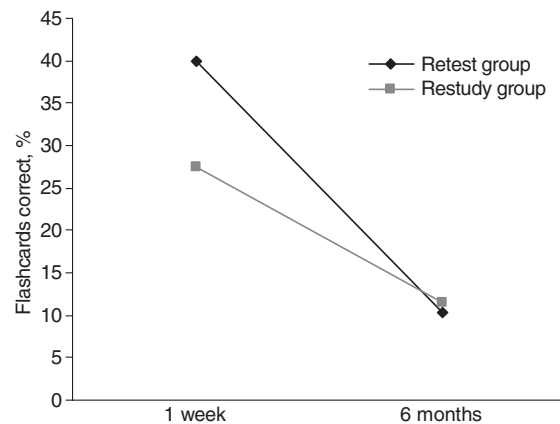

Figure 3 Effects of retesting and restudying on knowledge retention at 1 week and 6 months. Knowledge retention was assessed by asking participants to type in the target answers corresponding to the 30 flashcard cues. Flashcards for which all targets were given correctly, irrespective of their order, were considered 'as 'correctly recalled'

respectively. Retesting during a learning period of 1 day supports better knowledge retention over 1 week, but not over 6 months. The mean difficulty level (calculated by the percentage of participants who answered correctly) of the flashcards after 1 week was 0.62 (median = 0.62).

Metacognitive monitoring of medical students

To find out whether medical students were able to estimate the efficacy of their learning strategy, all participants were asked (Q2) after the learning phase to quantify how much the studying and testing periods had supported their learning by responding according to group to one of the items 'Repetitive studying was helpful for memorising the learning objectives' and 'Repetitive testing was helpful for memorising the learning objectives'. Responses were given on a 5-point Likert scale (1 = strongly agree, 5 = strongly disagree). No difference ($p = 0.15$) was detected between the groups (Fig. 4a). Additionally, students' judgement of learning (JoL) was assessed by asking the subjects (Q2) to predict how many flashcards they would be able to recall correctly after 1 week (Fig. 4b). Again, responses in the retest (77.9%) and restudy (74.5%) groups showed no significant difference. In order to make a historical comparison of medical students with non-medical students in terms of their metacognitive awareness, we asked a question published by Karpicke *et al.*[23] prior to our experiment: 'Imagine you are reading a textbook chapter for an upcoming examination. After you have read the chapter once, would you rather: (a) go back and restudy either the entire chapter or certain parts of the chapter, (b) try to recall material from the chapter (without the possibility of restudying the material), or (c) use some other study technique.' Table 2 shows that Munich medical students practised retesting (52%) more frequently than restudying (16%), and that this behaviour differs from the learning habits of US undergraduates. We asked an additional question previously used in 2007 by Kornell and Björk,[1] who surveyed college students about their study behaviours (Table 2). Like these college students, the vast majority of LMU Munich medical students practised retesting 'to figure out how well [they had] learned the information', not because they believed they might 'learn more that way'.

DISCUSSION

Factual knowledge is the basis for the complex intellectual competencies of doctors, such as clinical
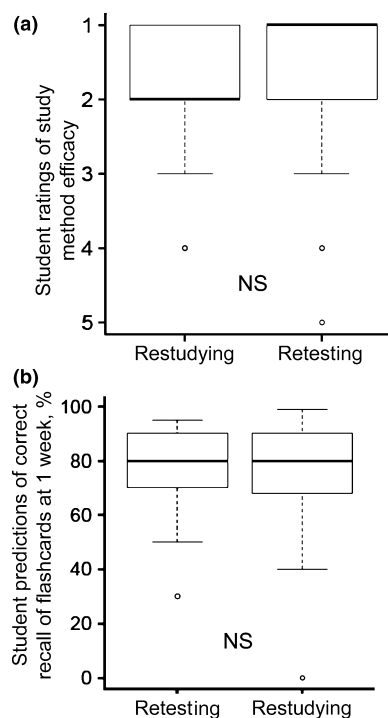

Figure 4 Metacognitive monitoring of medical students. Boxplots illustrate quartiles, extreme values and outliers. The bold line demonstrates the median. (a) Participant responses to the item 'Repetitive studying [or testing] was helpful in memorising the learning objectives', given on a Likert scale of 1–5 (1 = strongly agree, 5 = strongly disagree). (b) Participant predictions of the percentage of flashcards they would correctly recall at 1 week. NS = not significant

reasoning.[7–9] Modern curricula focus on instruction in practical skills and higher cognitive process dimensions such as the application or analysis of conceptual or procedural knowledge according to the revised Bloom's taxonomy.[30] Time spent on self-study is expected to support the acquisition of factual knowledge as it has been shown to be more efficient and to be more closely associated with improved achievement than time spent in lectures on the same topic.[31] Flashcards are frequently used by medical students for the purposes of memorising factual knowledge.[3,4] Most learners studying on their own drop flashcards they have memorised successfully from restudy in order to create more opportunities to study the remaining items.[32] As there is increasing evidence that retesting is more effective than restudying in terms of knowledge retention,[13,17,19,20] the online learning environment of our study was designed to investigate in a prospective, randomised trial whether successfully learned flashcards should preferably be dropped from either restudying or retesting protocols.

Table 2 Student responses to questions on metacognitive monitoring

Question	Medical students	Undergraduate students
After you have read the chapter once, would you rather:	$(n = 80)$*	$(n = 97)$[†]
(a) Go back and restudy either the entire chapter or certain parts of the chapter	16%	57%
(b) Try to recall material from the chapter (without the possibility of restudying the material)	52%	18%
(c) Use some other study technique	33%	21%
If you quiz yourself while you study, why do you do so?	$(n = 80)$*	$(n = 472)$[‡]
(a) I learn more that way than I would through rereading	16%	18%
(b) To figure out how well I have learned the information I am studying	65%	68%
(c) I find quizzing more enjoyable than rereading	3%	4%
(d) I usually do not quiz myself	16%	9%

* 80 medical undergraduate students at Ludwig-Maximilian University (LMU), Munich
[†] 97 undergraduate students at Washington University in St Louis (Karpicke et al.[23])
[‡] 472 undergraduate students at the University of California Los Angeles (Kornell and Björk[1])

Acquisition of knowledge

Time on task was similar in both groups, although time was not limited by the computer. The retest and restudy groups showed a difference of only 3% in the number of learning trials undertaken and overall time on task. We had previously conducted a pilot experiment, which led us to expect these results, but were still surprised as the active entering of answers using a keyboard obviously takes a substantial amount of time. Effective cognitive time on task may therefore have been even lower in the retest group compared with the restudy group. However, the degree of retrieval may be more intense when the participant is actively typing in the target answers.

The speed of learning and rate of success seen in A1 (administered after a distractor task at the end of day 1) were identical in both experiment groups, which leads us to conclude that retesting is not superior to restudying in terms of immediate knowledge acquisition.

Improved short-term retention

At the end of Week 1, we found a large relative difference of about 30% in the retention of knowledge and an absolute difference of 12% between the groups, representing a drop in the percentage of correctly recalled flashcards from 40% in the retest group to 28% in the restudy group, which corresponds to a high-intermediate to large effect. However, this difference is notably smaller than in the experiment by Karpicke and Roediger[17] (restudy group: 36%; retest group: 80%). This may reflect the stricter control of confounders in the latter study.[17] We intentionally chose an ecologically more valid learning domain in medical education. Our setting was more difficult than that used by Karpicke and Roediger[17] because it required the recall of not just one but up to five items in order to eliminate one flashcard. This explains the comparatively poor level of achievement in A2 and A3 with respect to the proportion of complete flashcards recalled correctly. The level of prior knowledge varied somewhat across our participants, but was never completely lacking, whereas all students in the study by Karpicke and Roediger[17] started from a position of no knowledge. Additionally, college students in an arbitrary sample are presumably less motivated to learn Swahili than medical students in the clinical phase of training are to learn nephrology. Although subjects had not been informed about the procedure, it is possible that some students may have expected an assessment at the end of the first week and had prepared themselves accordingly. Despite these open questions and possible confounders, we obtained highly significant results at the conclusion of Week 1 with a large effect size in favour of retesting versus restudying.

Lack of long-term effects

Quite recently, Larsen et al.[25] showed that retesting spaced over 4 weeks led to significantly better knowledge retention in medical residents compared with restudying after 6 months. We therefore re-assembled our participants at 6 months for the purposes of long-term follow-up. At this time-point, their correct recall of only 11% of the flashcards

represented a rather poor achievement, which, more surprisingly, did not differ between the two experiment groups. This result strongly supports the findings of others[32,33] and indicates that spacing out rather than massing study opportunities on a specific learning objective is more effective in terms of long-term retention. However, many students prefer to mass their study as they consider spaced practice to be less effective.[34] Although the practice of dropping items after a single correct recall has been shown to be ineffective, learners tend to do so when they are allowed to self-regulate their study time.[32] Thus, the use of strategies derived from evidence from cognitive psychology would seem to represent a reasonable way of ensuring that learning is most effective when using electronic flashcards as a learning modality.

Metacognitive self-monitoring

Metacognitive monitoring refers to an individual's understanding of the value of future study opportunities.[32] The accuracy of metacognitive monitoring is defined by its degree of resolution (high ratings for better-learned items) and calibration (the degree to which an individual's predicted recall levels match his or her actual recall levels). Poor calibration in terms of overconfidence may be one reason why self-directed learners increasingly drop items. Karpicke and Roediger[17] clearly showed that students using the more efficient retesting strategy predicted the same rate of recall (50% of the learning objectives) after Week 1 as the group using repetitive study. In our study students were more confident and predicted a recall rate of about 76% of learning cards after one week. Furthermore, Karpicke *et al*.[23] investigated whether students practise retrieval when they study on their own and found that 57.4% of their respondents claimed to practise restudying, but only 17.8% reported testing themselves after studying. This strongly contrasts with the claims of our medical students, who frequently (52%) practised retesting but rather rarely (16%) used restudying. These results support the hypothesis that medical students are trained in learning large amounts of information and may use retesting rather frequently in comparison with undergraduate students from other domains. We hypothesised that medical students lack awareness of the efficacy of different learning strategies. This was supported by our students' self-assessment at the end of the first experimental day, in which predictions for week 2 were similar in both experiment groups. The fact that testing without feedback promotes the retention of knowledge indicates that retrieval is itself a mnemonic enhancer. This finding accords with published data from non-medical cog-

nitive psychology.[35,36] In addition, retrieval has been shown to enhance encoding during a subsequent study period, a phenomenon described as the 'potentiating effect' in memory research.[37,38]

Limitations of the study and further questions

We present the first study in clinical medical students to compare the effectiveness of restudying and retesting strategies in the acquisition of relevant factual knowledge using a realistic online learning unit. However, the sample size was too small to allow a subgroup analysis to show whether the benefits are dependent on prior knowledge. Furthermore, the newly developed flashcard set should be evaluated in student cohorts from other medical schools in a multicentre study. We focused on specific content in nephrology for proof of principle, but follow-up experiments should address several areas of medical content. The comparison of our questionnaire results with historical data suggests that medical students differ from other students with respect to their learning strategies. This assumption should be evaluated in direct comparisons. In terms of the further development of retesting as a learning strategy in the actual medical curriculum, several questions remain unresolved. How many repetitions are required to optimise time efficacy? What is the optimal spacing of retesting? Is it advantageous to mix content from different content domains during the retesting sessions? Should the number and spacing of retesting sessions be adapted to the difficulty level of the flashcards?

Impact on the learning of factual knowledge

Most research into the role of retrieval in learning has used word pairs in paired-associate tasks. Very recently, these direct effects of test-enhanced learning have been shown to apply to the learning of text-based material.[39] Furthermore, it has been proposed that the repetition of retrieval does not need to be spaced over time to promote long-term retention.[39] However, the results of our study do not support this assumption. The responses to the questionnaires reveal that medical students are similarly insensitive to the efficacy of retrieval, as described elsewhere in the literature.[23] This has practical implications for self-regulated learning. The selection of learning strategies is strongly dependent on subjective assessment of current performance, albeit that such assessments may not correspond to actual future long-term performance.[36] It seems particularly important for future learning concepts to consider the evidence for the superiority of retesting over restudying because the use

of learning strategies can be actively guided during an online learning course. It has been suggested that test-enhanced learning can be implemented most efficiently by using active production tests that are spaced over time and for which feedback is given with some delay after the test.[12] Most investigations into test-enhanced learning have been conducted in laboratory settings and the need for further research in more realistic settings has been noted.[40] Our study strongly supports the importance of retesting in the retention of medical knowledge. However, it also shows the limitations of the method, particularly in terms of its lack of long-term effects.

REFERENCES

1 Kornell N, Björk RA. The promise and perils of self-regulated study. *Psychon Bull Rev* 2007;**14**:219–24.
2 Kornell N, Son LK. Learners' choices and beliefs about self-testing. *Memory* 2009;**17**:493–501.
3 Sleight DA, Mavis BE. Study skills and academic performance among second-year medical students in problem-based learning. *Med Educ Online* 2006;**11**:1–6.
4 Allen EB, Walls RT, Reilly FD. Effects of interactive instructional techniques in a web-based peripheral nervous system component for human anatomy. *Med Teach* 2008;**30**:40–7.
5 Bottiroli S, Dunlosky J, Guerini K, Cavallini E, Hertzog C. Does task affordance moderate age-related deficits in strategy production? *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2010;**17**:591–602.
6 Woods NN. Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med Educ* 2007;**41**:1173–7.
7 Groen GJ, Patel VL. Medical problem-solving: some questionable assumptions. *Med Educ* 1985;**19**:95–100.
8 Joseph GM, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. *Med Decis Making* 1990;**10**:31–46.
9 Patel VL, Kaufman DR. On poultry expertise, precocious kids, and diagnostic reasoning. *Acad Med* 1994;**69**:971–2.
10 Muijtjens AM, Hoogenboom RJ, Verwijnen GM, van der Vleuten CP. Relative or absolute standards in assessing medical knowledge using progress tests. *Adv Health Sci Educ Theory Pract* 1998;**3**:81–7.
11 McLachlan JC. The relationship between assessment and learning. *Med Educ* 2006;**40**:716–7.
12 Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. *Med Educ* 2008;**42**:959–66.
13 Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 2006;**17**:249–55.
14 McDaniel MA, Roediger HL III, McDermott KB. Generalising test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 2007;**14**:200–6.
15 Jacoby LL, Wahlheim CN, Coane JH. Test-enhanced learning of natural concepts: effects on recognition memory, classification, and metacognition. *J Exp Psychol Learn Mem Cogn* 2010;**36**:1441–51.
16 Galvagno SM Jr, Segal BS. Critical action procedures testing: a novel method for test-enhanced learning. *Med Educ* 2009;**43**:1182–7.
17 Karpicke JD, Roediger HL III. The critical importance of retrieval for learning. *Science* 2008;**319**:966–8.
18 Pyc MA, Rawson KA. Why testing improves memory: mediator effectiveness hypothesis. *Science* 2010;**330**:335.
19 Carrier M, Pashler H. The influence of retrieval on retention. *Mem Cognit* 1992;**20**:633–42.
20 Glover JA. The 'testing' phenomenon: not gone but nearly forgotten. *J Educ Psychol* 1989;**81**:392–9.
21 Roediger HL. The power of testing memory: basic research and implications for educational practice. *Psychol Sci* 2006;**17**:249–55.
22 Hogan RM, Kintsch W. Differential effects of study and test trials on long-term recognition and recall. *J Verb Learn Verb Behav* 1971;**10**:562–7.
23 Karpicke JD, Butler AC, Roediger HL III. Metacognitive strategies in student learning: do students practise retrieval when they study on their own? *Memory* 2009;**17**:471–9.
24 Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ* 2009;**43**:21–7.
25 Larsen DP, Butler AC, Roediger HL III. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 2009;**43**:1174–81.
26 McGuire FL. Psycho-social studies of medical students: a critical review. *J Med Educ* 1966;**41**:424–45.
27 Henning K, Ey S, Shaw D. Perfectionism, the impostor phenomenon and psychological adjustment in medi-

cal, dental, nursing and pharmacy students. *Med Educ* 1998;**32**:456–64.

28   Kopp V, Stark R, Fischer MR. Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. *Med Educ* 2008;**42**:823–9.

29   Simonsohn AB, Fischer MR. [Evaluation of a case-based computerised learning program (CASUS) for medical students during their clinical years]. *Dtsch Med Wochenschr* 2004;**129**:552–6.

30   Krathwohl DR. A revision of Bloom's taxonomy: an overview. *Theor Pract* 2002;**41**:212–8.

31   Gijselaers WH, Schmidt HG. Effects of quantity of instruction on time spent on learning and achievement. *Educ Res Eval* 1995;**1**:183–201.

32   Kornell N, Björk RA. Optimising self-regulated study: the benefits – and costs – of dropping flashcards. *Memory* 2008;**16**:125–36.

33   Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol Bull* 2006;**132**:354–80.

34   Simon DA, Björk RA. Metacognition in motor learning. *J Exp Psychol Learn Mem Cogn* 2001;**27**:907–12.

35   Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? *J Memory Lang* 2009;**60**:437–47.

36   Karpicke JD. Metacognitive control and strategy selection: deciding to practise retrieval during learning. *J Exp Psychol Gen* 2009;**138**:469–86.

37   Kornell N, Hays MJ, Björk RA. Unsuccessful retrieval attempts enhance subsequent learning. *J Exp Psychol Learn Mem Cogn* 2009;**35**:989–98.

38   Izawa C. The test trial potentiating model. *J Math Psychol* 1971;**8**:200–24.

39   Karpicke JD, Roediger HL III. Is expanding retrieval a superior method for learning text materials? *Mem Cognit* 2010;**38**:116–24.

40   Cantillon P. Do not adjust your set: the benefits and challenges of test-enhanced learning. *Med Educ* 2008;**42**:954–6.

APPENDIX 1
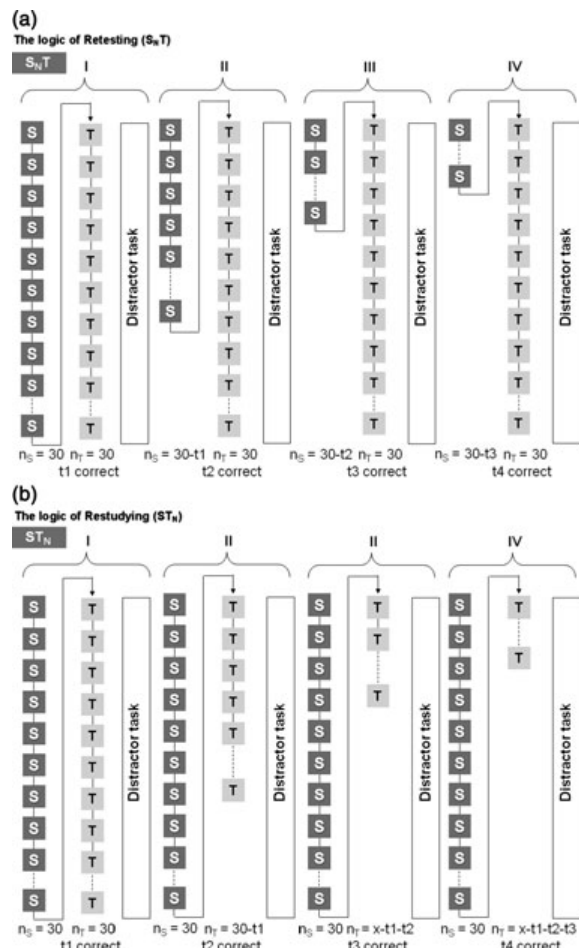
The logic of retesting and restudying.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article. Available at http://online library.wiley.com/doi/10.1111/j.1365-2923.2011.04043.x/suppinfo

Appendix S1. Medical content of flashcards.

Appendix S2. Standardised introduction for participants.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than for missing material) should be directed to the corresponding author for the article.