

# Unprepared Retesting of First Year Knowledge: How Much Do Second Year Medical Students Remember?

Margot M. Weggemans<sup>1</sup>  · Eugène J. F. M. Custers<sup>1</sup> · Olle Th. J. ten Cate<sup>1</sup>

Published online: 18 July 2017

© The Author(s) 2017. This article is an open access publication

## Abstract

**Introduction** With the implementation of integrated curricula, less time is spent on teaching basic sciences to the benefit of subjects with more clinical relevance. Even though learning in a clinical context seems to benefit medical students, concerns have been raised about the level of (bio)medical knowledge students possess when they enter their rotations. This study aimed to obtain empirical data on the level of knowledge retention of second year medical students at the University Medical Center Utrecht, the Netherlands.

**Method** A longitudinal study was performed in which second year medical students were retested for retention of first year knowledge by a study test consisting of questions from two course examinations of year 1, each with an interval of 8 to 10 months. Results were compared in a within-participants design.

**Results** The results of 37 students were analysed. Students scored on average 75% ( $\pm 8.2\%$ ) correct answers during the initial unit examinations and 42% ( $\pm 8.8\%$ ) for the knowledge retention test. With correction for guessing this was 71% ( $\pm 9.3\%$ ) versus 33% ( $\pm 9.9\%$ ), which means knowledge retention was on average 46%. Knowledge retention was higher for multiple choice questions (MCQs) (53%) versus non-MCQs (41%), and somewhat different for the two courses (53% and 40%).

**Conclusion** After an interval of 8–10 months, more than half of first year knowledge cannot be reproduced. Medical

students and faculty should be aware of this massive loss of knowledge and provide means to improve long-term retention.

**Keywords** Knowledge loss · Knowledge retention · Long-term retention · Retesting · Undergraduate medical education

## Introduction

Traditionally, the medical curriculum consisted of two phases: a preclinical phase in which the basic sciences were taught in the form of individual disciplines (such as anatomy, physiology and pathology) and a clinical phase in which students walked the wards. As the medical sciences—basic sciences as well as clinical sciences—have expanded and specialised considerably in the last century, medical faculties have struggled with the question which of this knowledge, and how much, to include in the core curriculum. Increasingly, stimulated by the movement of Problem Based Learning from the 1970s, the traditional discipline-based curricula have been replaced by curricula in which traditional basic science subjects are integrated with clinical subjects in many countries [1]. Today's medical curricula often consist of a sequence of units or blocks, such as “the Cell”, “Infection,” and “the Elderly Patient.” Basic science subjects that did not appear to directly contribute to the development of clinical competence have gradually been left out of the curriculum, to the benefit of subjects that have more relevance for clinical practice. In spite of this, pressure to add content to the core curriculum has remained high due to the inclusion of increasingly more clinical knowledge and the introduction of new skills to align the course with the CanMEDS roles. Inevitably, this has led to a considerable decrease of the time students spend learning

✉ Margot M. Weggemans  
m.m.weggemans@umcutrecht.nl

<sup>1</sup> University Medical Center Utrecht, Center for Research and Development of Education, Heidelberglaan 100, HB room 4.05, P.O. Box 85500, 3584 CX Utrecht, the Netherlands

basic science subjects, specifically in countries such as the Netherlands without a national examination focusing on the basic sciences [2].

There is some evidence that students in a curriculum featured by vertical integration and early clerkships suffer less from lack of knowledge [3–5] and that there is an advantage to learning relevant concepts in the context of clinical problems [6]. However, at least in the Netherlands, concerns have been raised that medical students do not possess sufficient readily available knowledge when they enter clinical rotations [7], even though all Dutch medical schools have moved towards more integrated curricula over the past decades. Due to the compressed curriculum, many topics are dealt with only once during undergraduate education, and this may be insufficient for students to establish a firm base of readily available knowledge. In any case, all other things being equal, students are less likely to recall knowledge they have been exposed to only once or twice than knowledge they have seen more frequently [8–10]. If no or very infrequent rehearsal is the norm for knowledge acquired in undergraduate medical education, a major loss of this knowledge before students enter the wards may not come as a surprise.

On the other hand, concerns about students appearing on the ward without sufficient knowledge are not new, far from that: they can be found in Cole [11], Dornhorst and Hunter [12], Neame [13], and Anderson [14], just to mention a few. Rarely, however, are such complaints supported by empirical evidence. For example, the strong claim that many students “retain a mere ten percent of the anatomy or biochemistry offered in the traditional first year course” [15] is not supported by evidence—even though such data appear to have been collected [16]. A review study by Custers [17] identified 20 empirical studies that investigated long-term retention of basic science knowledge in medical school. The results show a wide range of performances on tests for retention of this knowledge, from a level “not significantly different from what could be obtained by random guessing” (anatomy knowledge after 2 years [18]) to a comparatively marginal decrease from 72.6 to 69.7% correct answers on a 240-item multiple choice question (MCQ) test after an interval of 15 months [19]. Custers [17] gives rough estimates of 70% retention after 1 year, 40–50% retention after 2 years, and 30% retention after 4 years. More precise estimates are not possible, because these studies differ widely in coverage of basic sciences and type of questions in the tests, and many studies are incompletely reported, e.g., do not provide any information on possible rehearsal of the knowledge during the retention interval.

In short, the literature on retention of knowledge learned in medical school—and in academic courses in general—allows for two general conclusions. First, there is no consensus about how much knowledge students have lost when they enter the clinical clerkships. Second, *if* knowledge deficit is a problem, then one way to address this is by frequent rehearsal and

repeated testing during the retention interval, when students are attending courses on different subjects [20, 21].

At the University Medical Center Utrecht (UMC Utrecht) the preclinical curriculum, i.e. the first 2 years of a 6-year course, consists of units that are organised around specific themes or organ systems (such as *Circulation, Infection & Immunity, Metabolism, and Healthy & Diseased Cells*). Each unit extends over 5 or 6 weeks. The subject matter primarily consists of basic science and clinical knowledge, taught in an integrated fashion. Nevertheless, students’ performance on questions that involve basic sciences in the Utrecht Progress Test, administered in years 4 and 5—which purports to test for knowledge every medical doctor should possess at graduation—has been found wanting. To improve this situation, the faculty now considers introducing knowledge retention tests (called “CRUX-tests”) in the undergraduate curriculum. The idea is that in their second year, students will be retested for first year knowledge, and in the third year, for knowledge acquired in their second year. The assumption is that students will be aware of this future requirement and more actively rehearse and review the pertinent subject matter before taking these CRUX-tests and that as a consequence, they will show improved knowledge retention in the later years. However, before embarking upon such a curriculum innovation, we wanted to investigate the actual forgetting of knowledge in our undergraduate students. Therefore, we designed a study that directly probes long-term knowledge retention of first year basic sciences in UMC Utrecht medical students. We recruited volunteer undergraduate students to retake unit tests after an extended interval during which no formal education was scheduled on the topics of these units. The explicit aim was to quantitatively assess unrehearsed knowledge retention in these students. We selected two units, *Healthy & Diseased Cells* and *Metabolism*, as a compromise between scope (we wanted to capture a sufficiently broad domain) and feasibility (the practical impossibility of re-administering all unit tests). As the unit tests contain multiple choice questions as well as open-ended questions, we could also check for differences in long-term retention as expressed through these different types of items. Finally, we explored possible relationships between the students’ performance at the unit tests and their knowledge retention.

## Method

### Design

The design was a longitudinal study (repeated measures) in which first year medical students who made the original unit tests were retested at the beginning of their second year with the same tests. This study was approved by the Netherlands Association for Medical Education (NVMO) Ethical Review

Board. Students who volunteered to participate were retested for knowledge of the *Healthy & Diseased Cells* and *Metabolism* units which were given in the first months of the year 1 program (September through December 2013). Students were retested in September 2014; hence, the time lag between the course examinations in year 1 and the knowledge retention test in year 2 ranged from 8 to 10 months. This time lag was defined as the retention interval.

## Participants

All UMC Utrecht year 2 medical students were approached during the first block of their second year. Students were informed that they would have to complete a knowledge retention test, but were not informed on the nature of the knowledge they would be tested for. Forty-two students volunteered to participate in this study.

## Materials

As the original unit tests take approximately 2–3 hours to complete, we believed it would be too demanding for our students to require them to sit two complete tests. Hence, we split up the two unit tests in halves, and prepared four versions of a knowledge retention (KR) test by combining two test halves into one test form. Thus, each form of the KR test consisted of half of the questions of the test of *Healthy & Diseased Cells* and half of the questions of the test of *Metabolism* (Appendix). The main topics of these units were anatomy, biochemistry, cell biology, endocrinology and (patho)physiology. In the *Healthy & Diseased Cells* course students study normal growth and functioning of cells and tissues, and how defects in these processes can lead to diseases such as cancer and cystic fibrosis. In *Metabolism*, the anatomy and physiology of the gastrointestinal tract and the process of metabolism are taught, including the pathophysiology of obesity and related diseases. The unit test of *Healthy & Diseased Cells* was administered on November 7, 2013, the unit test of *Metabolism* on December 13, 2013. To ensure comparability of the test forms, the two test halves for each unit contained questions of approximately the same difficulty level (based on  $p$  values of the test items), with matched pairs of questions of similar  $p$ -levels being assigned randomly to the test halves. Open-ended questions—some of which contained several sub-questions—were assigned in alternation to the two test halves; these questions were not split up. The two halves of the *Healthy & Diseased Cells* test were labelled ‘A’ and ‘B’, the two halves of the *Metabolism* test were labelled ‘C’ and ‘D’. These four halves were then combined into four different forms: AC, AD, BC, and BD. The four test versions contained 49, 48, 49, and 48 questions, respectively. Each version consisted of 42–43 multiple choice questions, 1–2 “fill-in-the-blanks” and extended matching questions (FIB/EM

questions) and 3–6 open questions. Despite these differences, the four test versions can be considered equivalent (parallel) tests of the same subject matter, with approximately half of the test score (awarded points) being determined by the MCQ questions, and half by the other types of questions. The knowledge retention test was administered on paper, as was the original *Metabolism* unit examination. The *Healthy & Diseased Cells* original unit test was computer based.

## Procedure

Participants who volunteered were invited to make the test in a group session specially organised for this purpose on September 9, 2014. Twenty-nine students attended this session and delivered filled-in test forms. Individual appointments were made with 13 additional participants who were unable to attend this session. On October 14, 2014, data collection was completed. Students could take at maximum 2 hours to make the KR test; all participants managed to finish well within this time window. The four KR test versions were alternately distributed over participants in order of appearance, to ensure that we would end up with approximately the same number of completed test forms of each version. At the beginning of the session, students signed an informed consent form. By signing this form, they allowed the researchers to procure the corresponding examination results from the unit coordinators of the two units. Without these data, no longitudinal comparison would be possible. It was also pointed out to the students that the faculty would not be informed about their results on the KR test. After they completed and handed in the test form, they received a gift voucher worth €15 which they could spend in local shops. Finally, at the time the data were analysed and the results were available, all participants who indicated on the consent form they wanted to be informed about their results were sent an email with their results and the average scores of the whole group. All participants were informed that the KR test was voluntary and that scores obtained would not in any way have an influence on study results.

## Analysis

The answers to all questions were assessed in accordance with the assessment methods of the original unit examinations. All open-ended questions were checked using model answers by two of the authors independently (MMW and EJFMC). Discordance in judgement was resolved by consensus discussion. In a few cases, the unit coordinator was consulted on the assignment of points to a specific open-ended question. All correct answers to MCQs were awarded 1 point; fill-in-the-blanks questions were awarded 1 point for each correct (sub-)answer and the same holds for open-ended questions, though sometimes fractions of full points were awarded. Some

of the open-ended questions were scored like a spelling test, with points being detracted from a maximum obtainable score (awarded for a complete model answer) that were missing in the student's answer. Consequently, partly correct answers to open-ended questions could be awarded 0 points in case of too many omissions. For reasons of clarity and comparability between test versions, the total number of points obtained was expressed on a 0–100 percentage scale, i.e. a 100% score means the maximum number of obtainable points or 100% correct answers.

For every participant, a “virtual” unit examination test score was calculated by adding the points obtained on the two halves of the original unit test that corresponded with the two halves of the KR test (i.e. AC, AD, BC, or BD). Thus, the unit test scores were treated as if students had made the AC, AD, BC or BD test versions 8–10 months before even though they were never administered as such to the students (but they had seen each question at either the *Healthy & Diseased Cells* or the *Metabolism* examination). In the remainder of this article, we will refer to this “virtual” unit examination by ‘unit test’.

To enable comparison, the scores of these unit tests were also expressed as percentage of the maximum number of points that could be obtained. Longitudinal comparisons of the average scores for the unit test and the corresponding KR test (AC, AD, BC, or BD) were performed, for the test as a whole, as well as for two different subparts (MCQs versus FIB/EM and open-ended questions; and questions from *Healthy & Diseased Cells* versus *Metabolism*). Because of the small number of FIB/EM and open-ended questions, we only contrasted MCQs with other questions, i.e., non-MCQs. As our aim was to measure the extent of knowledge loss over the retention interval, Cohen's  $d$  as measure of effect size and the reliability interval of our measurements were our primary dependent measures. We also investigated whether knowledge retention was related to achievement on the unit examinations (i.e. whether “better” students remembered more).

## Results

Forty-two students volunteered to participate in this study and completed the KR test. Four students failed the first *Healthy & Diseased Cells* examination but successfully took the retest. The results of these four students were removed from final analysis, because they had restudied the material within the retention interval for a regular retest. Including their results would have inflated retention scores. The results of one other student were excluded because this participant had not taken the initial examination of *Healthy & Diseased Cells*. None of the participating students failed the *Metabolism* examination on the first occasion. All together, we included 10, 8, 8, and 11

completed forms of the AC, AD, BC, and BD test versions, respectively, in the analysis. Table 1 summarizes our results.

**Results of the Complete KR Test** The average score on the unit examination was 75% ( $\pm 8.2\%$ ) of the maximum obtainable points, against 42% ( $\pm 8.8\%$ ) on the KR test (effect size Cohen's  $d = 3.9$ ).<sup>1</sup> This difference is large enough to render a test for statistical significance superfluous, i.e. there is hardly any overlap between the two distributions. On average, students' scores decreased by 33% (percentage points) from the unit test to the KR test. Further analyses revealed a correlation of  $r = 0.57$  between students' scores on the unit test and the KR test, and a correlation of  $r = 0.40$  between students' scores on the unit test and knowledge loss (expressed as decrease in percentage points). This latter correlation means that students who scored high on the unit test lost more knowledge in *absolute* terms than students who scored low on the initial test. The correlation between students' scores on the unit test and their *proportional* knowledge retention was 0.04, i.e. the scores on the unit test did not predict at all what proportion of their knowledge students retained at the KR test. For example, a 50% proportional knowledge loss means a greater loss of knowledge in absolute terms for someone with a high score for the unit test compared to a student that scored low on the unit test.

**Results of the Complete KR Test with Correction for Guessing** As all test versions included a number of multiple choice questions (42 in test versions AC and AD, 43 in test versions BC and BD), we repeated the above analysis and applied a correction for guessing. To achieve this, for each participant, the expected total number of points that could be obtained by randomly choosing any of the alternatives in the MCQs was detracted from the actual number of points obtained. As this correction for guessing had to be applied to both the unit test and the KR test, it does not affect the size of the knowledge decrement expressed as percentage points. However, it enabled us to estimate the *relative* knowledge loss in our participants. The repeated analysis revealed an average of 71% ( $\pm 9.3\%$ ) of the maximum points that could be obtained at the unit test against 33% ( $\pm 9.9\%$ ) at the KR test. From this result, we can infer that students remember approximately 46% (95%-reliability interval of this estimate approximately 42%–50%) of the knowledge they acquired during the courses after a delay of approximately 9 months, under the assumption that they started the course with zero knowledge. Further analyses showed a correlation of  $r = 0.54$  between students' scores on the unit test and the KR test, a correlation of  $r = 0.42$  between students' score on the unit test and knowledge loss,

<sup>1</sup> All results are rounded to full percentages and to one decimal for standard deviations, in order to prevent an overestimation of exactness for these data.



**Table 1** Mean scores and effect sizes of the unit examinations and knowledge retention test with and without correction for guessing

	Complete test (without correction for guessing)	Complete test (corrected for guessing)	MCQs (corrected for guessing)	Non-MCQs	Healthy & Diseased Cells (corrected for guessing)	Metabolism (corrected for guessing)
Mean score unit examination (% ( $\pm$ SD))	75 ( $\pm$ 8.2)	71 ( $\pm$ 9.3)	75 ( $\pm$ 10.8)	69 ( $\pm$ 12.0)	68 ( $\pm$ 11.8)	75 ( $\pm$ 11.1)
Mean score KR test (% ( $\pm$ SD))	42 ( $\pm$ 8.8)	33 ( $\pm$ 9.9)	40 ( $\pm$ 14.7)	28 ( $\pm$ 9.3)	36 ( $\pm$ 11.7)	30 ( $\pm$ 12.2)
Effect size (difference unit examination— KR test) (Cohen's <i>d</i> )	3.9	4.0	2.7	3.8	2.7	3.8
Decrease in mean score (%)	33	38	35	41	32	45
Mean knowledge retention (%)	56	46	53	41	53	40

KR test knowledge retention test, MCQ multiple choice question

and  $r = 0.11$  between students' scores on the unit test and their proportional retention of knowledge.<sup>2</sup>

**Results of the Multiple Choice Questions** After correction for guessing, students scored on average 75% ( $\pm$ 10.8%) points on the MCQs of the unit test against 40% ( $\pm$ 14.7%) on the MCQs of the KR test (effect size Cohen's  $d = 2.7$ ). This suggests students retain approximately 53% (95%-reliability interval of this estimate approximately 48%–58%) of the knowledge they have acquired at the end of the course, as assessed by the results of MCQs.

**Results of the Non-MCQs** Non-MCQs can be of three types: fill-in-the-blanks, extended matching, or reasoning. For practical purposes, we considered the possibility of obtaining points by guessing the correct answer on these questions negligible. As of each type, only few questions were included, we lumped them together in a single analysis, which can be contrasted with the results of the MCQs. On the unit test, participants scored 69% ( $\pm$ 12.0%) of the points that could be obtained, against 28% ( $\pm$ 9.3%) at the knowledge retention test (effect size Cohen's  $d = 3.8$ ). If assessed by open-ended questions, students retain approximately 41% (95%-reliability interval of this estimate approximately 36%–46%) of the knowledge acquired as a consequence of attending the courses.

**Results of the Healthy & Diseased Cells Part of the KR Test (with Correction for Guessing)** On the set of test questions from the unit test for the *Healthy & Diseased Cells* unit, students scored 68% ( $\pm$ 11.8%) of obtainable points, against 36% ( $\pm$ 11.7%) on the knowledge retention test (effect size Cohen's  $d = 2.7$ ). Students retained approximately 53% (95%-reliability interval of this estimate approximately 48%–58%) of the knowledge acquired in the *Healthy & Diseased Cells* course.

<sup>2</sup> These correlations differ slightly from the correlations for the test without correction for guessing because the different test versions did not contain exactly the same number of MCQs and hence the number of points subtracted in the correction procedure differed slightly between test versions.

**Results of the Metabolism Part of the KR Test (with Correction for Guessing)** On the set of test questions from the unit test for the *Metabolism* unit, students scored 75% ( $\pm$ 11.1%) of obtainable points, against 30% ( $\pm$ 12.2%) on the knowledge retention test (effect size Cohen's  $d = 3.8$ ). Students retained approximately 40% (95%-reliability interval of this estimate approximately 35%–45%) of the knowledge acquired in the *Metabolism* course.

**Differential Decay of Healthy & Diseased Cells versus Metabolism Knowledge (with Correction for Guessing)** ANOVA revealed a significant interaction between knowledge domain and time of testing: knowledge loss was less for the *Healthy & Diseased Cells* course (from 68% to 36% percentage of obtainable points) than for the *Metabolism* course (from 75% to 30% percentage of obtainable points),  $F(1, 36) = 20.41$ ,  $p < 0.001$ . Thus, knowledge of *Healthy & Diseased Cells* was relatively better remembered after 8–10 months than knowledge of *Metabolism*.

## Discussion

In this study, we investigated knowledge retention in second year medical students by retesting them using the same tests (unit examinations) they had made approximately 8–10 months earlier. Assessed in this fashion, our results showed a massive loss of knowledge over the course of this retention interval. The effect sizes we found of the differences between scores on unit examinations or parts of unit examinations and the corresponding knowledge retention test or parts of this test ( $d$ -values ranging from 2.7 to 4.0), are exceptionally high for educational studies. Though providing exact quantitative values might not be justified given the limited accuracy of assessing knowledge by academic examinations, we feel safe to say that at best half of the knowledge acquired by our students in their first year courses, but probably even somewhat less, was retained after an interval of 8–10 months. This estimate may even be somewhat inflated should our assumption that students started the course with zero knowledge be

unjustified. It can also not be excluded that some students revisited some of the knowledge after passing the unit examinations, though it is unlikely that anyone did this to an extent that it might have seriously influenced his or her results on the knowledge retention test. In any case, our overall knowledge retention estimate of 56% (46% corrected for guessing) appears to be much lower than the 70% retention after 1 year reported in a review by Custers [17], though this review includes some studies that did yield results comparable to ours. For example, Arzi et al. [22] found approximately 53% retention of knowledge of the Periodic Table in high school students after a retention interval of 1 year. Sullivan et al. [23] report a reduction of approximately 51% of paediatric knowledge in final year medical students after a retention interval of 1 year. D'Eon [24] tested second year medical students after 10–11 months and found considerably more retention of physiology and immunology knowledge (80%) than neuroanatomy knowledge (47.5%). The difference appears to be accounted for by differential review or use during the retention interval: students revisit physiology and immunology subject matter, but not neuroanatomy (during the retention interval, no courses in neurology were scheduled).

Like neuroanatomy in the D'Eon [24] study, subject matter of the *Healthy & Diseased Cells* and *Metabolism* courses in our study was, in all likelihood, not revisited by students in our study during the retention interval. In addition, unit examinations include questions that do not test for core curricular knowledge that students are expected to remember—in detail—in the long run, but are primarily used to assess whether students sufficiently studied and mastered the subject matter dealt with in the unit. One unit coordinator (of the *Metabolism* course) explicitly said of some test questions that he did not expect students to remember the knowledge 1 year down the line. In addition, there appears to be a discrepancy between claims to include mostly or only knowledge that is relevant for future practice in the undergraduate courses, and traditional academic student assessment. If students are indeed expected to retain most or all of the knowledge they acquire in the units, standards for passing an academic test would have to be set much higher than is customary in “traditional” examinations. This implies that only questions with relatively high *p* values can be included in these examinations, which leaves little room for distinguishing average from excellent students.

The fact that some open-ended questions were scored in accordance with the “spelling test scoring” probably also has contributed to relatively low retention in our students. When comparing the results of the different question types, we found 53% knowledge retention for the MCQs (after correction for guessing), against 41% for the FIB/EM/reasoning questions. Since the same standards for the assessment of the KR test were used as for the original unit examinations, the same level of detail was expected in the answers given at the KR test in order to receive full credits for an answer. The answers

students gave to some of the open questions showed they remembered parts of the specific topic, but not enough to give an exact explanation or provide a fully detailed answer. For example, remembering the principle or working mechanism of a specific pathway but not being able to name specific receptors or transporters may have led to zero or only a small proportion of maximum attainable points being granted, when in fact some knowledge of the general concept was still available. This finding also illustrates that “knowing” something may not be a dichotomous, all-or-none affair, but may be partial, an issue that, to the best of our knowledge, is rarely discussed in the medical education literature.

In general, we found no correlation between students' scores on the unit test and *proportional* retention on the KR test. Though this might come as a surprise—intuitively, we might expect students who perform well on the examination to remember proportionally more than students who do not perform that well—the absence of such a relationship is actually in line with the literature [25–27], at least for noncumulative knowledge domains. It is the cumulative experience of many spaced learning episodes, rather than intense cramming for a test after a short course, that leads to knowledge in memory becoming stable and permanent, even though cramming may be necessary to obtain a high grade on a particular exam [28]. Or, to put it differently, students who attain high grades in initial tests do so because they remember more at the time of testing; students who attain high grades in cumulative tests, such as end-of-year examinations, do so because they know more, e.g. have durable knowledge [29].

We also obtained some results that we did not predict, and hence, want to discuss with caution. First, we found a significant difference in knowledge loss between the two units, *Healthy & Diseased Cells* and *Metabolism*. Scores on the unit test were higher for *Metabolism* than for *Healthy & Diseased Cells* (75% versus 68%), but this pattern was reversed on the retention test (30% versus 36%). In fact, this differential loss might even be underestimated, for some aspects of the design worked to favour retention of *Metabolism* knowledge. First, the unit test of *Healthy & Diseased Cells* was administered electronically (computer based) and second, during part of the unit test, students were allowed to consult their books, whereas the retention test was administered by paper forms and no books were allowed. Third, the retention interval for *Healthy & Diseased Cells* was longer by over 1 month than the retention interval for *Metabolism*. Nonetheless, students appeared to have better long-term memory for the former than for the latter knowledge. There is some anecdotal evidence suggesting that students may incidentally revisited knowledge of *Healthy & Diseased Cells* during their first year, whereas this is less likely for *Metabolism* knowledge.

One limitation of our study could be that students who participated were not representative of their class; that as a group, they would perform better than the average student.

This “volunteer effect,” i.e. overrepresentation of above-average students in educational studies, has been repeatedly reported [30–32]. In order to check for the presence of this effect, we compared our participants’ scores on the first two unit examinations with the scores of the remaining students who did this exam in their class. The results showed a clear volunteer effect. The average score of our participants for the unit examination of *Healthy & Diseased Cells* was 64% versus 57% for non-participants. For *Metabolism*, the average grades were 71% for study participants and 64% for non-participants. However, as we used within participant comparisons, and our primary aim was to measure knowledge retention rather than knowledge level, this volunteer effect cannot have affected the general results of our study. The finding of an essentially zero correlation between scores on the unit test and proportional knowledge loss strongly suggests that “poorer” and “better” students do not differ in this respect; hence, there is no reason to assume that our results cannot be generalized to the class as a whole. It is not impossible, on the other hand, that we slightly overestimated the general knowledge level.

An additional limitation could be that students were less motivated to perform well for our knowledge retention test as there was no reward for passing the exam nor a consequence of failing. We do however believe that the effects on our results are limited, as students could volunteer to participate in taking the KR test. Also, if motivation played a role, this would probably have had greater influence on the non-MCQs as they take greater effort to complete, but even for the MCQs a large loss of knowledge was found. Moreover, before and after completion of the KR test, students often mentioned that they were curious to know how much they remembered, which implies they made an effort to complete the test to the best of their abilities.

Another interesting question is whether students’ performance could have been inflated because they remembered the questions which they had seen less than 1 year before. The results of Sullivan et al. [23] suggest this does not play a major role: students in this study performed similar on questions they had seen before and new questions. Incidentally, our study reinforces this conclusion: in one extended matching question (from *Healthy & Diseased Cells*) students had to match five one-sentence descriptions of different diseases with the names of these diseases. By accident, the first five disease names in the list of 20 diseases matched the first five descriptions, in exactly the same order. Thus, by filling in “A, B, C, D, E,” for the first five diseases, the maximum number of points could be obtained. Surprisingly, not a single participant was aware of this; apparently, it did not occur, even to students who matched the first three descriptions, “A, B, C,” with the first three diseases, that proceeding with “D, E,” for the next diseases would be the correct answer. In other words, they had no memory of this question.

Finally, do our results support the faculty’s decision to introduce a new cycle of assessments, the “CRUX” tests, in order to boost retention of core knowledge (biomedical as well as clinical) acquired by students in the early years of their study? Our results show that less than half of first year knowledge is retained after an interval of less than 1 year, indicating that few students who are retested for first year knowledge in their second year would pass this test. An important difference between our study’s knowledge retention test and the CRUX-tests that will be implemented, however, is that students participating in our study were not informed on the content of the test, were therefore unable to prepare for the test and did not need to pass. Students who will make the CRUX-tests as part of their bachelor in the new curriculum will know exactly what the content of these tests will be and will be encouraged to rehearse this knowledge during the year and repeat the knowledge of the relevant units before entering the test. We believe that a repetitive, delayed testing of the same knowledge (in our case the CRUX-tests) will contribute to improved long-term retention of knowledge acquired in the undergraduate curriculum.

**Acknowledgements** The authors would like to thank Paul Steenberg and Marti Bierhuizen for providing the data of the unit tests, and for their time and additional information during the whole process. Without their cooperation this study could not have been performed. The authors also thank Joost Koedam, Annemarie Ultee-van Gessel, Inge van den Berg, Tom Roeling, and Carrie Chen for their time and helpful additions to the manuscript. Last but not least, the authors thank all 42 students courageous enough to expose their loss of knowledge to the benefit of medical education research.

#### Compliance with Ethical Standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

#### Appendix

Sample questions of the *Healthy & Diseased Cells* and *Metabolism* courses

(*p* values calculated exclusively for students who participated in the study)

##### Healthy & Diseased Cells—Question 1

If extracellular signals are absent, then what will most animal cells do?

Options: [a] they will stop their metabolism; [b] they will go into apoptosis; [c] they will go into a stage of rest; [d] they will decompose their cAMP

Correct answer: [b]

*p* value unit test: 0.94; *p* value knowledge retention test: 0.89

### Healthy & Diseased Cells—Question 2

Which of the following membrane specializations (junctions) provide mechanical solidity between two adjacent intestinal epithelial cells by connecting the intermediary filaments of these two cells?

Options: [a] desmosomes; [b] tight junctions; [c] gap junctions; [d] adherens junctions

Correct answer: [a]

*p* value unit test: 0.94; *p* value knowledge retention test: 0.39

### Metabolism—Question 1

Which of these organs is usually located entirely or partly in the retroperitoneal space?

Options: [a] spleen; [b] descending colon; [c] sigmoid colon; [d] stomach

Correct answer: [b]

*p* value unit test: 0.95; *p* value knowledge retention test: 0.37

### Metabolism—Question 2

If the mitochondria in a cell are partly “decoupled”, what will happen in this cell?

Options: [a] the TCA cycle will slow down; [b] the oxygen consumption per produced ATP will increase; [c] the electron transport chain will be slowed down; [d] electron transport will only be possible through FADH<sub>2</sub>

Correct answer: [b]

*p* value unit test: 0.42; *p* value knowledge retention test: 0.42

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Harden RM, Sowden S, Dunn W. Educational strategies in curriculum development: the SPICES model. *Med Educ.* 1984;18:284–97.
- Keijsers CJPW, Custers EJFM, ten Cate TJ. Een nieuw, probleemgeoriënteerd geneeskundecurriculum in Utrecht: minder kennis van de basisvakken. (A new, Problem Oriented Medicine curriculum in Utrecht: Less basic science knowledge). *Ned Tijdschr Geneesk.* 2009;153(34):400.
- Dahle LO, Brynhildsen J, Behrbohm Fallsberg M, Rundquist I, Hammar M. Pros and cons of vertical integration between clinical medicine and basic science within a problem-based undergraduate medical curriculum: examples and experiences from Linköping, Sweden. *Med Teach.* 2002;24(3):280–5.
- Kamalski DMA, ter Braak EWMT, ten Cate TJ, Borleffs JCC. Early clerkships. *Med Teach.* 2007;29(9):915–20.
- Wijnen-Meijer M, ten Cate O, van der Schaaf M, Harendza S. Graduates from vertically integrated curricula. *Clin Teach.* 2013;10(3):155–9.
- Norman GR, Schmidt HG. The psychological basis of problem-based learning: a review of the evidence. *Acad Med.* 1992;67(9):557–65.
- Hillen HFP (Ed.) (2012). *Geneeskunde Onderwijs in Nederland 2012. State of the Art Rapport en Benchmark Rapport van de Visitatiecommissie Geneeskunde 2011/2012.* (Medical Education in the Netherlands 2012. State of the Art Report). Quality Assurance Netherlands Universities (QANU), internal report, p. 1–66.
- Bahrick HP. Maintenance of knowledge: questions about memory we forgot to ask. *J Exp Psychol Gen.* 1979;108(3):296–308.
- Bahrick HP. Long-term maintenance of knowledge. In: Tulving E, Craik FIM, editors. *The Oxford handbook of memory.* New York: Oxford University Press; 2000. p. 347–62.
- Conway MA, Cohen G, Stanhope N. Very long-term memory for knowledge acquired at school and university. *Appl Cogn Psychol.* 1992a;6(6):467–82.
- Cole L. What is wrong with the medical curriculum? *Lancet.* 1932;220(5683):253–4.
- Dornhorst AC, Hunter A. Fallacies in medical education. *Lancet.* 1967;290(7517):666–7.
- Neame RL. The preclinical course of study: help or hindrance? *J Med Educ.* 1984;59(9):699–707.
- Anderson J. The continuum of medical education: the role of basic medical sciences. *J R Coll Physicians Lond.* 1993;27(4):405–7.
- Miller GE, Graser HP, Abrahamson S, Harnack RS, Cohen IS, Land A. *Teaching and learning in medical school.* Cambridge: Harvard University Press; 1961.
- Miller GE. An inquiry into medical teaching. *J Med Educ.* 1962;37(3):185–91.
- Custers EJFM. Long-term retention of basic science knowledge: a review study. *Adv Health Sci Educ.* 2010;15:109–28.
- Sinclair D. An experiment in the teaching of anatomy. *J Med Educ.* 1965;40:401–13.
- Swanson DB, Case SM, Luecht RM, Dillon GF. Retention of basic science information by fourth-year medical students. *Acad Med.* 1996;71:S80–2.
- Roediger HL 3rd, Karpicke JD. The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci.* 2006;1(3):181–210.
- Larsen DP, Butler AC, Roediger HL 3rd. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ.* 2009;43(12):1174–81.
- Arzi HJ, Ben-Zvi R, Ganiel U. Forgetting versus savings: the many facets of long-term retention. *Sci Educ.* 1986;70(2):171–88.
- Sullivan PB, Gregg N, Adams E, Rodgers C, Hull J. How much of the paediatric core curriculum do medical students remember? *Adv Health Sci Educ.* 2013;18(3):365–73.
- D’Eon MF. Knowledge loss of medical students on first year basic science courses at the University of Saskatchewan. *BMC Med Educ.* 2006;6:5.
- Bahrick HP. Stabilized memory of unrehearsed knowledge. *J Exp Psychol Gen.* 1992;121(1):112–3.
- Bahrick HP, Hall LK. Lifetime maintenance of high school mathematics content. *J Exp Psychol Gen.* 1991a;120:20–33.
- Conway MA, Cohen G, Stanhope N. On the very long-term retention of knowledge acquired through formal education: twelve years



- of cognitive psychology. *J Exp Psychol Gen.* 1991;120(4):395–409.
28. Conway M, Cohen G, Stanhope N. Why is it that university grades do not predict very long-term retention? *J Exp Psychol Gen.* 1992b;121(3):382–4.
  29. Conway MA, Gardiner JM, Perfect TJ, Anderson SJ, Cohen GM. Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *J Exp Psychol Gen.* 1997;126(4):393–413.
  30. Callahan CA, Hojat M, Gonella JS. Volunteer bias in medical education research: an empirical study of over three decades of longitudinal data. *Med Educ.* 2007;41:746–53.
  31. Rosenthal R, Rosnow R. The volunteer subject. In: Rosenthal R, Rosnow RL, editors. *Artifact in behavioral research.* NY: Academic Press; 1969. p. 59–118.
  32. Van Gelderen A, Ten Cate TJ. Vrijwilligerseffecten in tijdschrijfonderzoek. (Volunteer effects in time-on-task research). *Tijdschrift voor Onderwijsresearch.* 1985;10:149–60.