Open-book tests assessed

Heijne-Penninga, Marjolein

Publication date:
2010

6

Long-term effect of PBL and open-book assessment on knowledge retention

M. Heijne-Penninga, JBM Kuks, WHA Hofman, A Muijtjens, J Cohen-Schotanus
Submitted.

Abstract

The influence of problem-based learning (PBL) and open-book tests on long-term performance is unclear and has been subject of discussion for some time. With reference to the cognitive load theory the authors tested the hypotheses that PBL and open-book tests positively influence long-term knowledge retention.

Four progress test results of medical students of three different curricula were analyzed (n = 5478). The included curricula were: a PBL curriculum using closed-book tests for core knowledge and open-book tests to measure backup knowledge (PBLob), and a PBL and a traditional curriculum both using only closed-book tests (PBLcb and TC, respectively). All progress tests were divided into core and backup knowledge. Test results of fifth and sixth-year students were considered long-term performance. T-tests (with Bonferroni correction) were used to analyze differences between students of the three curricula.

In the long-term, PBL students performed significantly better than TC students on core knowledge (ES=0.35-0.86), PBLob students scored higher than PBLcb students (ES=0.36-0.62). TC students outperformed their peers in the first two study years (ES=0.28-0.91). First to third-year PBLob students performed worse than their peers on backup knowledge (ES=0.27-0.89), however, in the long term no differences in performance on backup knowledge were found.

In the long term, PBLob students scored best on core knowledge, followed by PBLcb. Referring to cognitive load theory it can be explained why PBL and open-book tests stimulate schemata construction, and why students are only able to profit from these schemata in the long-term.

Introduction

To function as medical professionals students must become active, independent learners and problem solvers who are able to use and manage knowledge, especially since the body of knowledge is growing and changing rapidly. The growth and depth of students' knowledge and, consequently, the ability to retrieve knowledge from memory is influenced by the instructional approach and the assessment programme.

Problem-based learning (PBL) is considered to be an instructional approach that suits the above-mentioned goals.[1,2] In a PBL curriculum, students deal with problems in small groups supervised by a tutor, and are assumed to be able to learn and recall information better than students in other kinds of curricula.[1,3] However, studies on the effectiveness of PBL have not revealed consistent outcomes, therefore the effectiveness of PBL is still subject of discussion.[4-7]

Assessment programmes also influence students' learning approaches and information processing strategies and, consequently, knowledge acquisition.[8,9] Open-book tests suit the need for processing the expanding body of knowledge.[10] During open-book tests students are allowed to consult references if they feel a need. The availability of reference sources during the tests influences the ways students handle knowledge.[11-13] The influence of open-book test preparation on long-term performance is not yet known.

In this study, we investigated the effect of a PBL approach, combined with open-book tests, on students' long-term knowledge retention. Referring to theories of memory and information processing, we expected that a combination of PBL and open-book tests yields better study results in the long term.

Memory and information processing

Knowledge must be stored in memory if it is to be retained and recalled in the long term. This memory storage process involves three successive stages: sensory memory, working memory and long-term memory.[14] *Sensory memory* receives input – presumably before it is recognized – and information is retained temporarily in the sensory registers. *Working memory* provides a means for using knowledge and is therefore the most active part of the memory system. The capacity of the working memory is limited: it can hold no more than five to nine elements of information and even fewer elements – possibly no more than two or three – can be processed simultaneously.[15-17] The *long-term memory* is the repository for more permanent knowledge and skills – information that is not currently being used but which is necessary for understanding.[18]

Several theories describe how information is stored in the long-term memory. The dominant view in psychology argues for the existence of a structuring mechanism – schemata, chunks et cetera – that determines the organization of knowledge so that it can be stored in and retrieved from memory efficiently. The more knowledge is embedded and connected with other information elements, the better it can be retrieved from long-term memory. Forgetting indicates a problem with retrieving information.[19]

A relevant theory for learning in a complex domain such as medical education is the cognitive load theory.[20] Medical curricula often involve authentic learning tasks which have many different solutions and a high degree of complexity. The cognitive load is critical when complex learning tasks are used, because these impose a high load on the learner's cognitive system. According to this theory, knowledge is stored in the long-term memory in schemata – cognitive structures that make up an individual's knowledge base.[21] When a schema is retrieved from the long-term memory by the working memory, it can be handled as a single element, irrespective of its complexity. Consequently, schemata not only organize knowledge but also reduce working memory load. Schemata are developed over a lifetime of learning and practice. They can be constructed by combining elements during the process of problem-solving, by incorporating elements in schemata that already exist in the long-term memory and by obtaining schematized information from others.[20]

PBL and open-book tests

Whether PBL promotes learning activities resulting in information storage in the long-term memory is subject of discussion. Kirschner et al. take the view that confronting students with complex tasks and letting them search for information and solve problems – as is the case in PBL – makes schemata construction more difficult.[5] They argue that too much working memory capacity is needed in problem-solving which means that it cannot be used for schemata construction. Schmidt et al. argue that PBL in particular allows for the flexible adaptation of guidance and may even suit a strategy in which cognitive structures are better organized.[6] An important aspect of PBL is the activation of prior knowledge upon which mental models are built. PBL stimulates elaboration on newly acquired knowledge, which also promotes the construction of cognitive schemata.[1,3] Moreover, PBL involves problem-solving in small groups, which stimulates schemata construction, in particular since the problems the students have to deal with are tailored to their knowledge levels.[20]

In general, studies concentrating on comparisons between the knowledge level of PBL students and students of traditional curricula are based on a (national) knowledge test for which students prepared extensively.[4,22-25] This might indicate that short-term performance in particular was assessed instead of long-term knowledge retention.

Assessment formats influence students' learning approaches and, consequently, knowledge storage in the long-term memory. When preparing for closed-book tests, students focus on short-term performance and in part will cram for a test; how they will perform in the long term remains uncertain.[26-29] The long-term effect of employing open-book tests has not yet been investigated. According to the cognitive load theory, medical students must incorporate new medical knowledge into schemata to be able to retrieve and process this knowledge in the future. When students are preparing for a closed-book test and are focused on the ability to recall knowledge during a one-time closed-book test to be taken shortly after learning, schemata building may occur sometimes but not as a matter of course. It is most likely that students will be more focused on remembering separate facts. When preparing for an open-book test students need to focus on the body of knowledge as a whole. The relationships between various pieces of information and where information can be found, are of particular importance. The ability to find and apply knowledge indicates a broader view of the learning content, which might result in a better way of structuring information to construct comprehensive mental schemata.[30]

In this study, the long-term performance of students participating in three different curricula was analyzed: a PBL curriculum using a combination of closed and open-book tests (PBLob), a PBL curriculum using only closed-book tests (PBLcb) and a traditional curriculum using only closed-book tests (TC). In the PBLob setting, closed-book tests were used to assess core knowledge and open-book tests were used to assess backup knowledge (knowledge students must be able to understand and apply with the help of references). With reference to the cognitive load theory we formulated the following hypotheses:

1. In the long term, students of both PBL curricula will perform better than TC students.
2. In the long term, PBLob students will perform better on backup knowledge than students of both other curricula.

Method

Participants and Context

Test scores of first through six-year students of three medical schools (n=5478) are included in this study. The first school has a problem-based curriculum and uses closed as well as open-book tests (PBLob, n=2192), the second school has a problem-based curriculum with only closed-book tests (PBLcb, n=1682) and the third school has a more traditional curriculum with closed book tests (TC, n=1,604).

All three medical schools offer a six-year medical programme. The Bachelor's programmes are knowledge-based and comprise pre-clinical training, whereas the Master's programmes are clinically oriented and involve several clerkships. The three programmes share the same learning objectives, which are set out in the National Blueprint for the Medical Curriculum.[31] Student populations of the schools are highly similar as a result of the entrance procedure. Admission to all Dutch medical schools is determined by a lottery system, weighted by grade point average, and students have limited influence on the specific university to which they are admitted.[22] The curricula use different teaching methods and assessment programmes to meet the final learning objectives.[32]

The PBLob curriculum involves many small group meetings in which the students discuss and solve patient problems. Relatively few lectures are given and the basic and clinical sciences are taught in an integrated way. The total body of knowledge is divided into core and backup knowledge. Core knowledge is assessed using closed-book questions. As students need to understand and apply backup knowledge, possibly with the use of reference sources, it is assessed using open-book questions. After the closed-book answers are handed in, the students are allowed to open their reference sources to complete the open-book part of the test.

The PBLcb curriculum is comparable to the above-mentioned PBLob curriculum, but relies on even fewer lectures. In addition, PBLcb students have more freedom to pursue their own learning goals. Knowledge is not divided into core and backup knowledge and written tests are solely of the closed-book type.

In the more traditional curriculum (TC) lectures are given on a regular basis and small group learning occurs infrequently. The basic and clinical sciences are partly integrated and students are not confronted with patient problems on a regular basis. Knowledge is not divided into core and backup knowledge and written assessment takes the form of closed-book tests only.

Progress test

In this study, knowledge was measured by the Progress test. Four times per year students of all participating Dutch universities simultaneously sit the same progress test. These tests aim to assess the students' functional knowledge – the knowledge that recently graduated medical doctors should be able to demonstrate immediately. The tests are largely curriculum-independent, with their content designed to comply with the Dutch National Blueprint for the Medical Curriculum.[22,31]

A single progress test consists of 200 multiple-choice questions constructed by the teaching faculty of the participating schools. Each test is jointly produced in accordance with the blueprint, which ensures stratification of the sample by discipline, disease or complaint categories. At each participating medical school the questions constructed by the school's faculty are reviewed by a committee. A question approved by the local review committee enters a central item bank and can be sampled for inclusion in a future progress test. Examination regulations ensure comparable test status across schools.[33,34] The information provided by these tests can help to identify differences in student performance between and within undergraduate medical curricula of the different schools.[35]

Procedure

Students' scores on the four progress tests of the academic year 2008–2009 were included in our study. Questions that were eliminated for statistical and content reasons by the committee responsible for the quality of the progress tests were excluded from our study. The curriculum coordinator (JBK), who is responsible for the content and construction of the entire PBLob curriculum, indicated for each question whether it focused on Bachelor's knowledge and whether it concerned core or backup knowledge. In addition, block coordinators who are responsible for the subsequent blocks, were asked to indicate which questions covered the theory of their block and whether they concerned core or backup knowledge. Questions that could not be classified were excluded from our study. This way, each progress test was divided into two subtests: core knowledge and backup knowledge. Test results of fifth and sixth-years students were considered long-term performance because of the time span between studying the knowledge and taking the test.

Analysis

For all subtests a percent-correct score was calculated for each cohort (year group) of each medical school. To avoid complications due to the interdependence of scores, the analysis was performed separately for each of the four progress tests. T-tests were used to compare percent-correct scores of all subtests. For each test moment, differences between the three schools (3 comparisons) for each of the six year groups and for two variables (core and backup knowledge) were analyzed, resulting in 36 comparisons. Due to the high number of comparisons we used

Bonferroni correction, with the Bonferroni-corrected alpha being 0.05/36, and therefore differences with $p \leq 0.0014$ were interpreted as significant. Furthermore, effect sizes (Cohen's d) were calculated to indicate the importance of the differences we found. An effect size of 0.80 was seen as large, 0.5 as medium and 0.2 as small.[36]

Results

The 4 subtests concerning backup knowledge included 109, 120, 120 and 122 questions respectively, whereas the 4 subtests concerning core knowledge consisted of 83, 65, 67 and 62 questions respectively. In total, 26 out of 774 questions (3.4%) concerned knowledge that was not explicitly dealt with during the Bachelor's programme of the PBLob curriculum. The percent-correct scores on the subtests are presented in Table 1.

Table 1. Mean scores per cohort, per medical school for each progress test

		Test 1 Sep 08		Test 2 Dec 08		Test 3 Feb 09		Test 4 May 09	
		Backup	Core	Backup	Core	Backup	Core	Backup	Core
Year 1	PBLob	6.21	12.11	6.94	14.98	10.69	12.42	12.51	21.30
	PBLcb	6.89	12.85	8.96	14.95	11.45	17.47	14.25	21.99
	TC	7.13	14.83	10.40	18.71	14.50	15.94	15.99	21.10
Year 2	PBLob	10.58	24.56	16.01	31.12	17.86	30.94	21.99	36.79
	PBLcb	10.46	21.41	17.49	25.63	22.20	24.11	25.49	35.26
	TC	14.78	28.63	22.36	30.27	23.66	32.09	30.04	34.57
Year 3	PBLob	19.33	40.43	25.67	40.43	26.36	40.32	31.50	46.14
	PBLcb	20.92	34.73	28.55	37.81	30.57	39.18	36.14	49.91
	TC	20.12	33.52	28.21	37.95	29.95	38.37	35.32	45.19
Year 4	PBLob	27.55	49.28	32.74	48.50	32.24	45.97	36.93	50.89
	PBLcb	25.50	44.15	33.89	43.61	35.39	45.05	38.63	51.38
	TC	24.57	42.67	34.24	44.20	35.18	46.28	37.22	48.86
Year 5	PBLob	35.03	58.37	39.51	56.21	39.13	55.12	43.06	57.79
	PBLcb	34.25	52.15	39.80	51.65	40.94	53.06	44.17	58.62
	TC	32.25	49.24	38.71	47.92	37.81	47.11	41.22	52.94
Year 6	PBLob	39.63	62.73	45.37	61.01	44.66	60.32	45.56	60.25
	PBLcb	40.59	60.33	46.11	57.54	45.33	57.13	46.15	60.92
	TC	38.40	55.75	44.96	53.84	43.32	54.44	44.70	57.37

Backup = the mean score on the subtest concerning backup knowledge, Core = the mean scores on the subtest concerning core knowledge

All students of the three medical schools performed better on core than on backup knowledge.

PBL

In the long term, PBLob and PBLcb students significantly outperformed TC students on core knowledge (ES=0.35-0.86). Only in years 1 and 2, TC students scored significantly higher on core knowledge than their peers in both PBL curricula (ES=0.28=0.91).

Table 2. Core knowledge: comparison of the mean scores of the three medical schools

		PBLob vs PBLcb			PBLob vs TC			PBLcb vs TC		
		diff		ES	diff		ES	diff		ES
Year 1	Test 1	-0.74		0.11	-2.71	*	0.40	-1.98	*	0.28
	Test 2	0.03		0.01	-3.74	*	0.63	-3.76	*	0.63
	Test 3	-5.05	*	0.84	-3.52	*	0.56	1.53		0.23
	Test 4	-0.69		0.09	0.19		0.02	0.89		0.11
Year 2	Test 1	3.15	*	0.44	-4.07	*	0.54	-7.22	*	0.91
	Test 2	5.48	*	0.65	0.84		0.10	-4.64	*	0.61
	Test 3	6.83	*	0.72	-1.15		0.12	-7.98	*	0.91
	Test 4	1.53		0.13	2.22		0.18	0.69		0.06
Year 3	Test 1	5.70	*	0.58	6.90	*	0.67	1.21		0.13
	Test 2	2.60	*	0.26	2.49		0.24	-0.14		0.01
	Test 3	1.14		0.11	1.95		0.18	0.81		0.08
	Test 4	-3.77	*	0.31	0.95		0.07	4.72	*	0.41
Year 4	Test 1	5.13	*	0.50	6.61	*	0.64	1.48		0.15
	Test 2	4.89	*	0.49	4.30	*	0.43	-0.59		0.06
	Test 3	0.91		0.08	-0.31		0.03	-1.23		0.12
	Test 4	-0.49		0.04	2.03		0.17	2.52		0.23
Year 5	Test 1	6.22	*	0.62	9.13	*	0.86	2.91		0.28
	Test 2	4.56	*	0.47	8.28	*	0.85	3.72	*	0.36
	Test 3	2.07		0.20	8.02	*	0.76	5.95	*	0.56
	Test 4	-0.83		0.08	4.85	*	0.45	5.68	*	0.54
Year 6	Test 1	2.40		0.28	6.98	*	0.69	4.58	*	0.45
	Test 2	3.47	*	0.39	7.17	*	0.75	3.70	*	0.40
	Test 3	3.19	*	0.36	5.88	*	0.59	2.69		0.28
	Test 4	-0.67		0.07	2.89		0.28	3.50	*	0.35

* $P \leq 0.0014$

Open-book tests

In the long term we found no differences between the scores of students of the three curricula on backup knowledge. However, fifth and sixth-year PBLob students outperformed their peers several times on core knowledge (ES=0.36-0.86). In the short term, the PBLob students scored significantly lower on backup knowledge than their peers (ES=0.27-0.89) especially compared with the TC students (Table 3).

Table 3. Backup knowledge: comparison of the mean scores of the three medical schools

		PBLob vs PBLcb			PBLob vs TC			PBLcb vs TC		
		diff		ES	diff		ES	diff		ES
Year 1	Test 1	-0.68		0.14	-0.92		0.19	-0.24		0.04
	Test 2	-2.03	*	0.43	-3.46	*	0.65	-1.44		0.23
	Test 3	-0.76		0.17	-3.81	*	0.76	-3.05	*	0.57
	Test 4	-1.74	*	0.31	-3.47	*	0.58	-1.74	*	0.29
Year 2	Test 1	0.12		0.02	-4.20	*	0.75	-4.32	*	0.71
	Test 2	-1.48		0.21	-6.35	*	0.88	-4.87	*	0.67
	Test 3	-4.34	*	0.71	-5.79	*	0.89	-1.45		0.21
	Test 4	-3.50	*	0.40	-8.05	*	0.89	-4.55	*	0.48
Year 3	Test 1	-1.58		0.19	-0.78		0.10	0.80		0.10
	Test 2	-2.88	*	0.32	-2.54	*	0.27	0.34		0.04
	Test 3	-4.22	*	0.48	-3.60	*	0.39	0.62		0.07
	Test 4	-4.64	*	0.45	-3.82	*	0.36	0.82		0.08
Year 4	Test 1	2.05		0.22	2.98	*	0.32	0.94		0.11
	Test 2	-1.15		0.12	-1.50		0.16	-0.35		0.04
	Test 3	-3.15	*	0.37	-2.94	*	0.35	0.21		0.02
	Test 4	-1.70		0.18	-0.29		0.03	1.41		0.15
Year 5	Test 1	0.79		0.08	2.79		0.28	2.00		0.21
	Test 2	-0.29		0.03	0.80		0.08	1.09		0.12
	Test 3	-1.81		0.21	1.32		0.15	3.13	*	0.34
	Test 4	-1.12		0.12	1.83		0.20	2.95		0.32
Year 6	Test 1	-0.97		0.10	1.23		0.13	2.20		0.22
	Test 2	-0.74		0.08	0.41		0.04	1.14		0.12
	Test 3	-0.67		0.09	1.34		0.15	2.01		0.24
	Test 4	-0.58		0.07	0.86		0.10	1.44		0.17

* $P \leq 0.0014$

Discussion

In this study we examined students' long-term performance on progress tests. The results showed that PBL students performed significantly better than TC students on core knowledge in the long-term, although TC students outperformed their peers in the first two study years. The use of open-book tests negatively influenced short-term performance. In the long term, no influence of open-book tests on the backup knowledge score was discerned and PBLob students performed better than their peers on core knowledge.

The hypothesis about PBL was confirmed by our results: in the long-term, PBL students outperformed their TC peers, however in the first two study years the outcome was the other way around. It is possible that the view of Kirschner et. al is right and that the short-term working memory of the PBL students might have been overloaded in the first two study years, resulting in more difficulties with short-term knowledge retrieval.[5] Probably, this is a problem with retrieval cues, which activate elements in the long-term memory so they can be retrieved. People can have difficulties activating a memory if they have been exposed to the wrong cues.[37] The TC students concentrated more on isolated knowledge and were therefore better able to retrieve that knowledge.

In the long term, PBL students outperformed TC students. Fifth and sixth-year students deal with knowledge within the context of their clerkships. Their experiences probably resulted in automated schemata that could be activated by a variety of cues and were less dependent on cues related to single situations or cases. Probably, PBL students had constructed schemata in their long-term memories better than TC students, which is in line with Schmidt's view, and in the long term they were able to retrieve them more easily.[7]

The hypothesis that in the long term PBLob students will perform better on backup knowledge than students of both other curricula was not confirmed by our results. No differences were found on long-term performance. In the short term PBLob students scored *lower* on backup knowledge. According to interference theory, the act of recalling specific information sometimes interferes with the retrieval of all of the original information.[38-40] For example, the act of remembering a couple of items on a longer list decreases the probability of remembering the other items on that list.[40] This might also be the case with recalling core and backup knowledge. Students need to remember the core knowledge as it is assessed using closed-book tests. In the short term, the retrieval of core knowledge probably interferes with the retrieval of backup knowledge. In the long term, the negative influence of open-book tests on backup knowledge performance disappeared. It is likely that fifth and sixth-year students are simultaneously exposed to core and backup knowledge during their clerkships. Such an exposure results in more complex schemata, in which core and backup knowledge are entirely integrated and become fully automated due to experience and practice.[20] This reduces the interference between these two kinds of knowledge and results in better backup knowledge retrieval and especially better core knowledge retrieval.

A strength of this study is that the test scores of a large number of students of three different curricula were included. In addition, even though the students from the three curricula were enrolled in different programmes, they were highly comparable due to the admission procedure.[22] Inter-curricular comparisons face some potential confounding issues and sources of error. It is impossible to maintaining blinding and most medical students have all the prerequisite skills needed for successful performance despite their curriculum.[4,41,42] To prevent these biases we used a curriculum-independent test as an outcome measurement, which concentrates on final objectives that are the same for all three curricula.[31] Moreover, all three medical schools participated in the test item construction, as advised by Muijtjens et al.,[34] and we also included several test measurements. Under these conditions, test results can be used to compare student achievement across different curricula.[34]

A possible limitation of this study is that core and backup knowledge were determined by one medical school. Teachers of other medical schools might have different opinions about core knowledge. However, all students performed much better on core than on backup knowledge, which indicates that the distinction between core and backup knowledge seems relevant. It is remarkable that two-thirds of the progress test questions concerned backup knowledge since the progress tests aim to assess mastery of the functional knowledge that every graduating physician should possess. Therefore, the question arises whether they actually assess the most relevant and required knowledge.

In conclusion, long-term performance was particularly positively influenced by a problem-based curriculum format. Using open-book tests next to closed-book tests did not influence long-term performance on backup knowledge, but did seem to improve performance on core knowledge. Referring to the cognitive load theory, it seems likely that in particular the problem-based instructional approach but also open-book tests stimulate schemata construction when storing knowledge in the long-term memory. Whereas students have difficulties retrieving these schemata in the short term, in the long term these difficulties seem to disappear and students profit from these schemata.

Acknowledgments

We are grateful to Mrs J. Bouwkamp-Timmer for her critical and constructive comments on several drafts of the manuscript.

References

1. Dolmans D, Schmidt H. The advantages of problem-based curricula. *Postgrad Med J* 1996;72:535–8.
2. Barrows HS, Tamblyn RM. Problem-based learning: an approach to medical education. New York: Springer 1980.
3. Schmidt HG. Foundation of problem-based learning: some explanatory notes. *Med Educ* 1993;27:422–32.
4. Colliver JA. Effectiveness of problem-based learning curricula: research and theory. *Acad Med* 2000;75:259–66.
5. Norman GR, Schmidt HG. Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Med Educ* 2000;34:721–8.
6. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experimental, and inquiry-based teaching. *Educ Psychol* 2006;41:75–86.

7. Schmidt HG, Loyens SMM, van Gog T, Paas F. Problem-based learning is compatible with human cognitive architecture: commentary on Kirschner, Sweller and Clark (2006). *Educ Psychol* 2007;42:91–7.
8. Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programs. *Med Educ* 2005;39:309–17.
9. Cohen-Schotanus J. Student assessment and examination rules. *Med Teach* 1999;21:318–21.
10. Heijne-Penninga M, Kuks JBM, Schönrock-Adema J, Snijders TAB, Cohen-Schotanus J. Open-book tests to complement assessment-programmes: analysis of open and closed-book tests. *Adv Health Sci Educ* 2008;13:263–73.
11. Broyles IL, Cyr PR, Korsen N. Open book tests: assessment of academic learning in clerkships. *Med Teach* 2005;5:456–62.
12. Theophilides C, Koutselini M. Study behavior in the closed-book and the open-book test: a comparative analysis. *Educ Res Eval* 2000;6:379–93.
13. Heijne-Penninga M, Kuks JBM, Hofman WHA, Cohen-Schotanus J. Influence of open and closed-book tests on medical students' learning approaches. *Med Educ* 2008;42:967–74.
14. Atkinson RC, Shiffrin RM. Human memory: a proposed system and its control processes. In: Spence KW, Spence JT, eds. *The psychology of learning and motivation: advances in research and theory.* New York: Academic Press 1968;89–195.
15. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956;63:81–97.
16. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. *Educ Psychol* 2002;38:1–4.
17. Kirschner PA. Cognitive load theory: implications of cognitive load theory on the design of learning. *Learn Instruc* 2002;12:1–10.
18. Bower GH. Cognitive psychology: an introduction. In: Estes WK, ed. *Handbook of learning and cognitive processes, Vol. 1. Introduction to concept and issues.* Hillsdale, NJ: Erlbaum 1975:25–80.
19. Byrnes JP. Memory. In: Byrnes JP. *Cognitive development and learning in instructional contexts*. Needham Heights, Massachusetts: Allyn & Bacon 1996:35–59.
20. Van Merrienboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;44:85–93.
21. Sweller J. Cognitive load during problem solving: effects on learning. *Cog Sci* 1988;12:257–85.
22. Schuwirth L, Bosman G., Henning RH, Rinkel R, Wenink ACG. Collaboration on progress testing in medical schools in the Netherlands. *Med Teach* 2010;32:476–79.
23. Dochy F, Segers M, Bosche PV, Gijbels D. Effects of PBL: a meta-analysis. *Learn Instruc* 2003;13:533–68.
24. Hoffman K, Hosokawa M, Blake R, Headrick L, Johnson G. Problem-based learning outcomes: ten years of experience at the university of Missouri – Columbio School of Medicine. *Acad Med* 2006;81:617–25.
25. Distlehorst LH, Dawson E, Robbs RS, Barrows HS. Problem-based learning outcomes: the glass half-full. *Acad Med* 2005;80:294–99.

26. Chan WP. Assessment of medical students' knowledge retention in a diagnostic radiology course: lecture attendees versus absentees. *Ann Acad Med Singapore* 2009;38:237–39.
27. Kerfoot BP, DeWolf WC, Masser BA, Church PA, Federman DD. Spaced education improves the retention of clinical knowledge by medical students: a randomized controlled trial. *Med Educ* 2007;41:23–31.
28. Ling Y, Swanson DB, Holtzman K, Bucak SD. Retention of basic science information by senior medical students. *Acad Med* 2008;83:82–5.
29. Custers EJFM. Long-term retention of basic science knowledge: a review study. *Adv Health Sci Educ* 2010;15:109–28.
30. Driscoll M. Psychology of Learning for Instruction, 3rd edn. New York: Allyn & Bacon 2005.
31. Van Herwaarden, CLA, Laan RFJM, Leunissen RRM. Raamplan artsopleiding 2009 [Dutch Blueprint for the Medical curriculum 2009]. Utrecht: Nederlandse Federatie van Universitaire Medische Centra 2009.
32. Schmidt HG, Cohen-Schotanus J, Arends LR. Impact of problem-based, active learning on graduation rates for 10 generations of Dutch medical students. *Med Educ* 2009;43:211–8.
33. Muijtjens AMM, Schuwirth LWT, Cohen-Schotanus J, Thoben AJNM, van der Vleuten CPM. Benchmarking by cross-institutional comparison of student achievement in a progress test. *Med Educ* 2008;42:82–8.
34. Muijtjens AMM, Schuwirth LWT, Cohen-Schotanus J, van der Vleuten CPM. Origin bias of test items compromises the validity and fairness of curriculum comparisons *Med Educ* 2007;41:1217–23.
35. Muijtjens AMM, Schuwirth LWT, Cohen-Schotanus J, van der Vleuten CPM. Differences in knowledge development exposed by Multi-curricular progress data. *Adv Health Sci Educ* 2008;13:593–605.
36. Hojat M, Xu G. A visitor's guide to effect sizes. Statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ* 2004;9:241–9.
37. Anderson JR. Learning and Memory: an integrated approach. New York: Wiley 1995.
38. Tomlinson TD, Huber DE, Rieth CA, Davelaar EJ. An interference account of cue-independent forgetting in the no-think paradigm. *PNAS* 2009;106:15588–93.
39. Tulving E, Arbuckle TY. Input and output interference in short-term associative memory. *J Exp Psychol* 1966;72:145–50.
40. Roediger HL, Schmidt SR. Output interference in the recall of categorized and paired associative lists. *J Exp Psychol Hum Learn Mem* 1980;6:91–105.
41. Norman G. RCT = results confounded and trivial: the perils of grand educational experiments. *Med Educ* 2003;37:582–4.
42. Schmidt HG. Innovative and conventional curricula compared: what can be said about their effects? In: Nooman ZM, Schmidt HG, Ezzat ES, eds. *Innovation in medical education: an evaluation of its present status.* New York: Springer Publishing Co. 1990;1–7.