

The effect of distributed practice on students' conceptual understanding of statistics

Luc Budé · Tjaart Imbos · Margaretha W. van de Wiel ·
Martijn P. Berger

Published online: 2 October 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In this study the effect of the reduced distribution of study activities on students' conceptual understanding of statistics is investigated in a quasi-experiment. Conceptual understanding depends on coherent and error free knowledge structures. Students need time to construct such knowledge structures. A curriculum reform at our university resulted in statistics courses which were considerably shortened in time, thereby limiting students' possibility to distribute study activities. Independent samples of students from before and after the reform were compared. To gauge conceptual understanding of statistics, students answered open ended questions in which they were asked to explain and relate important statistical concepts. It was shown that the reduction of distributed practice had a negative effect on students' understanding. The finding that condensed courses make it more difficult for students to reach proper understanding of the subject matter is of interest for anyone who is engaged in reforming curricula or designing courses.

Keywords Statistics education · Distributed practice · Spacing effect · Knowledge acquisition · Conceptual understanding · Problem based learning

Introduction

In this paper the conceptual understanding of topics taught in statistics courses is investigated and compared across course lengths. A number of characteristics of conceptual understanding are reported in the literature. Conceptual understanding enables the explanation of causal mechanisms and processes (Graesser et al. 2001; Noordman and Vonk 1998). It leads to better performance (Mayer 1989) and better application and transfer of what has been learned (Novak 2002; Feltovich et al. 1993). Conceptual understanding creates a sensation of coherence of the subject matter (Entwistle 1995). This implies that stored knowledge elements are not isolated or arbitrarily connected. Conceptual understanding of studied material is only possible if knowledge elements are stored in coherent

L. Budé (✉) · T. Imbos · M. W. van de Wiel · M. P. Berger
Maastricht University, Maastricht, The Netherlands
e-mail: Luc.Bude@MaastrichtUniversity.nl

structures that contain the relevant concepts of the domain, as well as their relations (Chi et al. 1981; Wyman and Randel 1998; Kintsch 1998). Our definition of an individuals' understanding is based on these findings. Conceptual understanding is shown, when a person demonstrates coherent, error free knowledge structures. In this view conceptual understanding of statistics is related to the quality of the knowledge structures of an individual learner.

For example, the appropriate choice of statistical techniques in a study and correct interpretation of the results requires conceptual understanding of the material. In the case of a *t*-test, this means that one has to know why this test should not be performed on categorical data, whether the data are independent or not, how sample size is related to the power of the test, how a *p*-value should be interpreted in the light of a possible type I error, how the *p*-value and effect size are related to the relevancy of the findings, and what hypothesis can be tested. If an individual demonstrates correct knowledge of how these matters are interrelated, then we assume that person has conceptual understanding of the subject matter. Only being able to reproduce a definition, or literally quote a passage from a textbook, or mastering a skill without knowing why and when it is applicable, does not demonstrate conceptual understanding.

Conceptual understanding can be measured through the use of questions that require the combination of distinct concepts (Olson and Biolsi 1991; McNamara et al. 1996). This also holds for assessing statistical understanding (Garfield 2003; Schau and Mattern 1997). Open ended questions that ask for explanations, and underlying causal mechanisms are well suited for this purpose (Budé et al. 2009; Feltovich et al. 1993; Gijbels et al. 2005; Jonassen et al. 1993). For example, the question how the width of a confidence interval is reduced by an increase of the sample size concerns a causal relationship. The answer requires a correct combination of the two concepts sample size and confidence interval. Answers to such questions can reveal how an individual links concepts together. This discloses the knowledge structures of an individual and thus enables the assessment of conceptual understanding as we have defined it.

The development of conceptual understanding, i.e. the construction of coherent, error free knowledge structures, is influenced by several factors. In this study we focused on the effect of distributed practice. Distributed practice is defined as study activities with intervals (e.g., Bahrick and Hall 2005). Two conditions were compared. In the first condition students followed a statistics course that lasted for 6 months. In the second condition, because of a curriculum change, the students studied the same subject matter in a course with a time span of only 8 weeks. The study activities of the students in the first condition were therefore more distributed.

Distributed practice versus massed practice

A change in the curriculum programming at our university enabled us to compare two different statistics courses. Before the curriculum change, statistics was taught in courses with a time span of 6 months. The statistics courses included lectures, problem based learning (PBL) group meetings with a tutor in which subject matter was discussed, and practical sessions to train the students in using statistical software. Between these educational activities the students had at least 1 week for self-study. The intervals of 1 week or more between educational activities and the possibility to spread self-study over several days created so-called distributed practice. So, students following this course, i.e. students

in the distributed practice condition, could spread their study activities over a time span of 6 months, with intervals that varied from several days to more than 1 week.

After the curriculum change, the statistics course was concentrated in a period of 8 weeks. As before the change, these condensed courses consisted of lectures, tutorial group meetings where the subject matter was discussed, and practical sessions. The lectures and practical sessions before and after the curriculum change were identical, the same books and literature were used, the PBL group meetings were similar, the number of educational activities was equal, students were instructed to study the same material during self-study, and the courses covered the same topics. The introductory courses, in which this study was carried out, covered the central limit theorem, sampling distributions, *t*-tests, analysis of variance (ANOVA), linear regression analysis, and methodological subjects such as randomised clinical trials and quasi experimental designs. So, it can be concluded that the courses before and after the change were comparable regarding content and implementation, and the total amount of time the students were supposed to spend on studying the subject matter.

The most important difference was that after the change the courses were much shorter and contained at least three educational activities each week, i.e. a lecture, a group meeting and a practical session. This reduced considerably both the spacing of these educational activities and the opportunity to spread self-study. Both the shorter intervals between educational activities and the reduced distribution of self-study led to more massed practice, or what Seabrook et al. (2005) have called clustered presentations of subject matter. Massed practice is defined as studying subject matter uninterruptedly or with only short breaks, in a brief time interval (e.g., Bahrlick and Hall 2005). Students following the course after the curriculum change will be referred to as students in the massed practice condition.

Beneficial effects of distributed practice in comparison to massed practice, i.e. the spacing effects, were found in research that at first predominantly was done in the laboratory and usually was focused at memorising lists of items with intervals in terms of minutes or hours (Glenberg 1979; Melton 1970). In subsequent work the positive effect of distributed practice was also established in authentic educational settings (Seabrook et al. 2005) and much longer time intervals (Bahrlick and Hall 2005; Bahrlick and Phelps 1987). Bahrlick and Hall explain the spacing effect with students' metacognitive monitoring. That is, when study activities are spaced in time, students will notice that they have forgotten some of the material in the intervening period. When they experience such retrieval failures, they will be inclined to use encoding methods that will lead to better retention. They may also select and plan study strategies that will lead to less forgetting (Benjamin and Bird 2006).

Based on this research concerning the distribution of practice it was expected that the massed practice that resulted from the curriculum change, would have a detrimental effect on the development of students' conceptual understanding of the subject matter. The research question of the study was: will the reduction of distributed practice lead to less conceptual understanding of the subject matter of an introductory statistics course.

This question led to three hypotheses:

1. Distributed practice will lead to better conceptual understanding compared to massed practice.
2. Conceptual understanding will in both conditions be better after the course than during the course. Students are expected to refine their knowledge structures during the course, because the subject matter is repeatedly dealt with in the course. Therefore, it is expected that in both the distributed practice condition as in the massed practice condition the knowledge structures will be more coherent, and error free at the end of the course.

3. The difference in conceptual understanding of the students in the distributed practice condition and the massed practice condition will be larger after the courses than during the courses. The effect of distributed practice will be the larger at the end of the course, because students will have had the largest amount of practice then. Thus, there will be an interaction between the kind of practice and the time factor.

Method

Design

Data were collected from independent samples of students in the population cohorts before (*Condition*_{6months}) and after the curriculum change (*Condition*_{8weeks}). All students were tested with the same open ended questions to measure conceptual understanding. This enabled a straightforward comparison between the conditions. The questions were directed at statistical hypothesis testing theory. Measurements were taken at two time points in both conditions: one group of students was measured during the course and another group directly after the course. The measurement during the course was immediately after the subject hypothesis testing was dealt with in the courses. It should be noted that recall and memory effects were controlled by independent sampling and random assignment of the students to different samples at the two time points in both conditions. This allowed the comparison of four independent groups of students with respect to their conceptual understanding of statistical hypothesis testing theory.

It was expected that students in the *Condition*_{6months} would do better on the open ended questions, thus showing better conceptual understanding. We also wanted to know whether this understanding would be relatively stable over a longer period. Therefore, we measured with the same open ended questions conceptual understanding of a fifth group of students from the *Condition*_{6months}, half a year after the course was concluded. For practical reasons it was not possible to do a comparable long term measurement for students in the *Condition*_{8weeks}.

The entire cohorts were compared with respect to the average grades on the final course exam and on the exam of another statistics course. This last comparison was done to uncover possible differences in the overall abilities of the two cohorts. Finally, members of the Department of Methodology and Statistics of Maastricht University were included in the study as an expert control group.

Participants

Participants were first year bachelor students of the Faculty of Health Sciences of our university. They were recruited from two cohorts of students in consecutive academic years, directly before and after the change in the curriculum programming. During recruitment the students were told that they had to answer questions about statistics and that they would be paid 10 euro. This payment was granted to avoid attracting only motivated students who were particularly interested in statistics. Only students who volunteered to participate were collected. This resulted in two pools of volunteers; one for the *Condition*_{6months} and one for the *Condition*_{8weeks}. All students took the introductory statistics courses in which this study was executed.

For both conditions the students were at random sampled from the two pools of volunteers and randomly assigned to the two measurement time points: during and directly after the course. In total 152 students participated. In *Condition*_{6months} 64 students participated, of which 33 during the course ($n_{1\text{during}} = 33$), and 31 after the course ($n_{1\text{after}} = 31$). In *Condition*_{8weeks} 58 students participated, of which 30 during the course ($n_{2\text{during}} = 30$), and 28 after the course ($n_{2\text{after}} = 28$). Finally, an independent sample of 30 students from the *Condition*_{6months} participated in the long term retention test half a year after the course ($n_{1\text{longterm}} = 30$). One-hundred-and-thirteen of all participants were female, 39 were male. Approximately 75 percent of the Health Sciences' students is female and so the sample was in proportion to the gender ratio in the larger population. The age of the participants ranged from 20 to 26 years. A separate group of nine faculty members of the Department of Methodology and Statistics ($N_{\text{M\&S}} = 9$) participated as expert control group. Their age ranged from 25 to 60 years and their teaching experience ranged from 4 to 30 years.

Measurement instrument

To assess conceptual understanding of the subject matter, ten open ended questions concerning statistical hypothesis testing were designed. These questions specifically asked for the explanations of and the relations between statistical concepts. For example, students were asked to explain the relation between the central limit theorem and the confidence interval. For a complete list of the questions see the “Appendix”. A detailed and elaborated answer key with thesaurus was formulated with the aid of four statisticians. The answers were split up in propositions and awarded with a weighted scoring system. For reliability purposes a second rater scored a random subset of the data ($N = 14$). Interrater agreement was high. The correlation between the scores was significant ($r = .950$), the mean of the two raters did not significantly differ ($M_A = 20.07$; $M_B = 18.36$; $t = 1.20$, $p = .250$), nor did the standard deviation ($SD_A = 15.6$; $SD_B = 16.7$).

The total number of points that could be obtained was 110. This number could only be obtained if all the details of the quite comprehensive answer key were mentioned in the answers. An upper score limit was determined by the answers of a sample of nine faculty members of the Department of Methodology and Statistics, who were considered to master the subject matter competently. Their average score served as a reference point.

In addition, students' final course exam grades were used. The grades served as an additional measure to test for the difference between the two courses. The exams for both cohorts of students consisted of 30 multiple choice questions. Both exams measured students' achievement toward the subject matter.

Procedure

Students received written and oral instructions to answer the questions as completely as possible, even if they doubted whether an answer was correct. This was done to get information on partial knowledge and possible misconceptions. Answers had to be written on a blank sheet of paper. Time was not limited; the whole procedure took on average 1 h.

Analysis

The written answers were copied for scoring purposes. The answers to the questions were awarded with points in accordance with the scoring system. A two (*condition*) \times two (*time*

point) fixed factor Analysis of Variance (ANOVA) was performed with the score of the open ended questions as dependent variable. Long term retention was assessed 6 months after the course was finished and compared with a *t*-test to the group of students from the *Condition*_{6months} that was measured directly after the course. The average grades of the two cohorts on the final course exam and on the exam of another statistics course were also compared with *t*-tests.

Results

Conceptual understanding test

The two (*condition*) \times two (*time point*) fixed factor Analysis of Variance (ANOVA) showed no significant interaction. The model without the interaction term showed significant differences of both conditions ($F(1,119) = 36.05, p < .000$) and time points ($F(1,119) = 19.07, p < .000$). Both models are presented in Table 1.

The long term retention measurement showed that the level of conceptual understanding decreased significantly over a period of 6 months ($t = 6.68, p < .000$) for the *Condition*_{6months}. The mean of the faculty members differed significantly from all four groups ($p < .000$). The descriptive statistics of the measurements are presented in Table 2. The overall means for the conditions were $M_{6months} = 30.8$ and $M_{8weeks} = 19.6$ (main effect condition). The overall means for the time points were $M_{during\ the\ course} = 21.5$ and $M_{after\ the\ course} = 29.7$ (main effect time point).

Table 1 Analysis of variance for the effect of condition and time point on conceptual understanding

Source	<i>df</i>	<i>F</i>	<i>p</i>
Model with interaction			
Time point	1	18.722	.000
Condition	1	35.908	.000
Time point \times condition	1	.153	.697
Error	118		
Model without interaction			
Time point	1	19.071	.000
Condition	1	36.049	.000
Error	119		

Table 2 Descriptive statistics of students' conceptual understanding

	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>	<i>N</i>
<i>Condition</i> _{6months} during the course	26.5	10.9	7.0	54.0	33
<i>Condition</i> _{6months} after the course	35.3	10.4	12.0	61.0	31
<i>Condition</i> _{8weeks} during the course	16.1	7.9	1.0	29.0	30
<i>Condition</i> _{8weeks} after the course	23.4	11.2	4.0	46.0	28
Faculty members	51.2	19.2	17.0	73.0	9
Long term retention (<i>Condition</i> _{6months})	14.73	7.74	2.0	27.0	30

Measurements are independent, from random samples

Exam grades

The mean grades on the final course exams of the entire cohorts were compared with a *t*-test. The results showed a significant difference between the two conditions ($M_{6\text{months}} = 6.45$, $M_{8\text{weeks}} = 5.9$, $t = 3.34$, $p < .001$). The mean grades on the course exam of another statistics course showed no difference between the cohorts ($M_1 = 5.8$, $M_2 = 5.7$, $t = .932$, $p = .352$).

Discussion

The first hypothesis, i.e., distributed practice will lead to better conceptual understanding compared to massed practice, was confirmed. During and immediately after the course the students in the distributed practice condition performed better. The change in the curriculum resulted in a considerable reduction of the time span of the statistics courses. The reduction of the distribution of the subject matter over time decreased the possibility of effectively using distributed practice. We found that the students in the 6 months group performed better than those in the 8 week group. Looking at the simple effects we found that students measured during the course in *Condition*_{6months} (distributed practice condition) scored significantly higher than the students measured during the course in *Condition*_{8weeks} (massed practice condition) and even better than those in *Condition*_{8weeks} after the course. That is, after the latter students studied the subject matter in preparation for the course exam. The students measured in *Condition*_{6months} after the course scored best. These results seem to be somewhat in contrast to the results of Donovan and Radosevich (1999). They found that the effect of distributed practice declined as task complexity increased. However, our results clearly show that the loss of the spacing effect negatively affected students' conceptual understanding of the course material. We believe that the statistical subject matter is rather complex for first year university students. The results of our study are more in line with the results of Rohrer and Taylor (2006). They found a beneficial effect of distributed practice on solving abstract mathematic problems. Our conclusion is supported by the grades on the final course exams. The mean of the whole cohort before the curriculum change altogether was significantly higher than the mean of the cohort after that change. Both exams were supposed to assess the same level of understanding of the subject matter.

The second hypothesis that conceptual understanding of the subject matter in both cohorts of students will be better after the course than during the course was also confirmed by our results. This growth in conceptual understanding indicates that the knowledge structures of the students gradually became more coherent, and error free during the course.

The third hypothesis, i.e. the difference in conceptual understanding between the students in the distributed practice course and the massed practice course will be larger after the courses than during the courses, was not confirmed. Students in both conditions showed a comparable improvement from within the course to directly after it. Despite the fact that the interaction was not significant, it was in the expected direction. The improvement of conceptual understanding over the course was somewhat less in *Condition*_{8weeks} than in *Condition*_{6months}.

The scores of the faculty members were taken as an upper bound. Faculty members answered the questions unprepared. In contrast, the students answered the questions after they had studied the subject matter. However, at the end of the introductory courses

students in both conditions, as expected, clearly performed below the level of the faculty members.

We compared both cohorts with respect to their grades on the final exams of another statistics course. We did not find a significant difference. This indicated that these cohorts were comparable. Our results, therefore, can be attributed to a positive effect of distributed practice. This corroborates previous research (Bahrck and Hall 2005; Bahrck and Phelps 1987; Seabrook et al. 2005).

Several conclusions can be drawn from this study. Firstly, distributed practice enhances short term conceptual understanding of statistics. During and directly after the course students in the distributed practice condition outperformed the students in the massed practice condition. Secondly, the level of conceptual understanding of the students was far below the expert level of understanding. It was not expected that after an introductory statistics course the students would be able to reach an expert level. In fact, this cannot be expected after any introductory course, regardless of the content. However, the difference between the scores of the experts and the students in both conditions was rather large, indicating a modest level of students' conceptual understanding. Moreover, it was shown in the long term measurement that this level of conceptual understanding declined substantially over a period of 6 months for students in the *Condition*_{6months}. We do not have long term retention data for the *Condition*_{8weeks}, but we assume that long term retention for this condition even would have been worse, because students in this condition already had less good results immediately after the course. Moreover, the benefit of distributed practice tends to increase with delay (Seabrook et al. 2005). So, probably students do not recollect much of what they have learned in statistics courses after a period of 6 months or more, especially after 'cramming' the subject matter. Yet, in subsequent statistics courses the subject matter of previous courses is assumed to be known. Student's more advanced learning depends on their retention of previous learned material (Willingham 2002). Without this assumed adequate prior knowledge it will be impossible for students to advance to higher levels of conceptual understanding. It seems to be advisable, therefore, to try to improve students' knowledge even after the statistics courses are finished. Distributed practice might be a fruitful approach in this respect. It could be demanded from students to apply in subsequent non-statistical courses what they have learned in the statistics courses. For example, when discussing scientific literature, the statistical analyses of the studies and the results could be scrutinised. This would cause the subject matter of the statistics courses to recur regularly. It has been shown that without a periodic recurrence of the subject matter, the relations between concepts will fade and conceptual understanding will decline (Conway et al. 1991, 1992; Neisser 1984; Semb and Ellis 1994). Future research is needed to study if and how this can be implemented and whether such an approach would lead to the anticipated positive effect. An appropriate methodology for such research might be design-based research, in which systematically adjusted aspects of education can be tested in naturalistic contexts (van den Akker et al. 2006; Barab and Squire 2004).

The instrument for the measurement of conceptual understanding consisted of ten open ended questions. The use of these questions, which asked for explanations and the relation between statistical concepts, was based on literature on comprehension (Olson and Biolsi 1991; McNamara et al. 1996), comprehension of statistics (Budé et al. 2009; Feltovich et al. 1993; Garfield 2003; Schau and Mattern 1997; Gijbels et al. 2005; Jonassen et al. 1993), and on our definition of conceptual understanding. The answers were scored on the basis of a comprehensive answer key. This answer key contained all elements of the possible answers down to the smallest detail. It was not expected that it would be possible to answer the questions so completely that the maximum of 110 points could be obtained.

Although time was not limited, the available time appeared to be insufficient to reproduce all minutiae. The answers of the faculty members therefore served as an upper limit. Scoring the answers to the open ended questions with the answer key, in combination with this reference point seemed to be an appropriate measurement of students' conceptual understanding.

The study presented in this paper was executed in an educational field setting. This might limit the scope of the conclusions somewhat, because we could not control all threats to internal validity. For example, it could be questioned, whether students in the distributed practice condition actually have studied regularly during the course, because we did not measure self-study. We are confident, however, that students did practice in a distributed way, because attending the meetings, in which exercises were discussed, were obligatory for the students. Attending these meetings every 2 weeks and discussing the subject matter created the distributed practice. Therefore, we are confident that our findings reflect genuine effects. Moreover, the students were randomly sampled from the volunteer pools and assigned to the measurement time points. Furthermore, as the studied conditions were embedded in authentic learning settings, it may be expected that the external validity of the outcomes is relatively high.

The finding that condensed courses make it difficult for students to reach proper conceptual understanding indicates that understanding of the subject matter could be improved if courses would be designed over a longer period of time.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix: list of questions

1. *Explain* the relation between the central limit theorem and the confidence interval. First *explain* both terms and then how they are related.
2. Suppose that a 95% confidence interval for μ is 85 to 105. Can it be assumed then, with 95% confidence, that μ can range between 85 and 105? *Explain* your answer.
3. *Explain* the formula: $\bar{x} \pm z_{1-\alpha/2}(\sigma/\sqrt{n})$. First *clarify* the different terms and then the formula (i.e. *explain* why the terms are in the formula).

Suppose a researcher studies the IQ of students. The mean IQ of the Dutch population is known to be 100. The hypotheses of the study are: $H_0: \mu = 100$; $H_1: \mu > 100$. A random sample is drawn from the population of students. The sample mean is 120.

4. *Explain* in your own words which procedure is needed in the example above to be able to reject or not reject the null hypothesis. *Explain* why.
5. Suppose the researcher, in the example above, decides to reject the null hypothesis. Is it certain then, that the alternative hypothesis is true? *Explain* your answer.
6. Suppose another researcher uses the same data as in the example above. This researcher does not reject the null hypothesis (without making a mistake). *Explain* what he may have done differently.
7. Suppose a researcher reports that the sample mean does not exactly match the hypothesised population mean. *Explain* why it is more informative to report the confidence interval and what extra information it provides.
8. *Explain* what a type II error is and when the probability of such an error increases.

9. *Explain* in your own words what a test statistic is and how it is used in tests.
10. When an absolute z -value increases, then the corresponding p -value decreases. *Explain* why, using the concept of normal distribution.

References

- Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, *52*, 566–577.
- Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over eight years. *Journal of Experimental Psychology*, *13*, 344–349.
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*, 1–14.
- Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language*, *55*, 126–137.
- Budé, L. M., van de Wiel, W. J., Imbos, T. J., Schmidt, H. G., & Berger, M. P. F. (2009). *A procedure for measuring understanding* (Submitted).
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.
- Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: Twelve years of cognitive psychology. *Journal of Experimental Psychology*, *120*, 395–409.
- Conway, M. A., Cohen, G., & Stanhope, N. (1992). Very long-term retention of knowledge acquired at school and university. *Applied Cognitive Psychology*, *6*, 467–482.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*, 795–805.
- Entwistle, N. (1995). Frameworks for understanding as experienced in essay writing and in preparing for examinations. *Educational Psychology*, *30*, 47–54.
- Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale: Lawrence Erlbaum Associates, Publishers.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22–38.
- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*, 27–61.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetition on recall and recognition. *Memory & Cognition*, *7*, 95–112.
- Graesser, A. C., Olde, B., & Lu, S. (2001). *Question-driven explanatory reasoning about devices that malfunction*. Paper presented at the Annual Meeting of the American Educational Research Association 2001. Seattle, April 10–14, 2001.
- Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale: Lawrence Erlbaum Associates, Publishers.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Mayer, R. E. (1989). Models for understanding. *Review of Educational Research*, *59*, 43–64.
- McNamara, D. S., Kintsch, E., Butler-Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596–606.
- Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting. *Journal of Experimental Psychology*, *113*, 32–35.
- Noordman, L. G. M., & Vonk, W. (1998). Memory-based processing in understanding causal information. *Discourse Processes*, *26*, 191–212.
- Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, *86*, 548–571.
- Olson, J., & Biolsi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 240–285). Cambridge, UK: Cambridge University Press.

- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209–1224.
- Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education*. Amsterdam: IOS Press.
- Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107–122.
- Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research, 64*, 253–286.
- van den Akker, J., Gravemeijer, K., McKenny, S., & Nieveen, N. (2006). *Educational design research*. London: Routledge.
- Willingham, D. T. (2002). How we learn. Ask the cognitive scientist: Allocating student study time. Massed practice versus distributed practice. *American Educator, 26*, 37–39.
- Wyman, B. G., & Randel, J. M. (1998). The relation of knowledge organization to performance of a complex cognitive task. *Applied Cognitive Psychology, 12*, 251–264.