# How Fast Do Students Forget What They Learn in Consumer Behavior? A Longitudinal Study

Donald R. Bacon and Kim A. Stewart

*The retention curve for knowledge acquired in a consumer behavior course is explored in a longitudinal study, tracking individual students from 8 to 101 weeks following course completion. Rasch measurement is used to link tests and to achieve intervally scaled measures of knowledge. The findings indicate that most of the knowledge gained in the course is lost within 2 years. Evidence is provided that knowledge acquired at a deep level of understanding is more likely to be retained than knowledge acquired at a surface level of understanding, and knowledge tested more than once during a course is more likely to be retained than knowledge tested only once. No significant differences in retention were observed related to material covered in a project. Implications are discussed.*

***Keywords:*** *very long-term memory; retention curve; deep learning; testing; Rasch measurement*

As educators, we would like to believe that the knowledge students gain in our courses is stored away for later use as their careers develop. For example, we often teach students how to make decisions as marketing managers. But what if the knowledge they seemed to have mastered on the final exam is not so permanently stored so that by the time a student needs some particular knowledge it is no longer available from memory? The purpose of the present research is to explore how well marketing knowledge is retained over the weeks and years after completing one particular marketing course, consumer behavior. After estimating the general shape of the retention curve, we will test three specific hypotheses concerning the effects of deep learning, repeated testing, and project-related learning on knowledge retention.

## BACKGROUND LITERATURE

The present research integrates two issues related to very long-term memory. The first and central issue is the estimation of how much learning is retained as time passes. Second, this research investigates selected factors hypothesized to increase long-term retention. The relevant literature in each of these areas is reviewed before discussing the methodology in detail.

### The Retention of Learning

The field of memory research can be divided into three broad categories concerning the retention interval—the length of time between when the original learning is completed and when the retention of that learning is assessed. Short-term memory research generally involves retention intervals measured in seconds. Long-term memory research examines retention intervals measured in minutes, days, or sometimes weeks. The term *very long-term memory* refers to studies of retention intervals of a few weeks to many years. The works of Harry Bahrick and his associates are often cited in the very long-term memory literature. For example, Bahrick, Bahrick, and Wittlinger (1975) found a somewhat surprising level of retention of names and faces over retention intervals that ranged from 2 weeks to 57 years. The present research focuses on the very long-term retention of consumer behavior knowledge, with retention intervals ranging from 8 to 101 weeks.

In another study, Bahrick (1984) found a surprising level of retention for knowledge of Spanish learned in college, a content area that is closer to the content of interest in the present study. Although knowledge of Spanish declined exponentially for the first 3 to 6 years, retention appeared to stabilize thereafter, not diminishing again until approximately

Don Bacon is an associate professor of marketing in the Daniels College of Business at the University of Denver. Kim Stewart is an independent researcher.

The financial support received from the Daniels College of Business for this research is gratefully acknowledged.

*Journal of Marketing Education*, Vol. 28 No. 3, December 2006 181-192
DOI: 10.1177/0273475306291463

30 years later. This later forgetting was partly attributed to the onset of senility among some of the study participants. Based on these observations, Bahrick posited the existence of a "permastore" and asserted that knowledge that becomes part of the permastore is essentially impervious to forgetting over practical time periods. Although we might take heart from this study and hope that our teachings enter students' permastores, this outcome is unlikely because of the limited exposure most students have to consumer behavior. Bahrick found that college students who had the equivalent of only one course in Spanish often lost all of their Spanish knowledge within a few years. Only those with more extensive training exhibited retention consistent with the permastore concept (some had the equivalent of 10 college courses).

A set of studies quite closely related to the present research was reviewed by Semb and Ellis (1994). The 62 studies in Semb and Ellis's meta-analysis include research of disciplines as diverse as language, psychology, engineering, and biology but not business education. Based on their meta-analysis, Semb and Ellis concluded that students retain a substantial amount of knowledge learned in school. However, their use of ordinal measures in the relative loss function used to evaluate retention raises doubts about their conclusions. Semb and Ellis defined the measure taken at the end of a particular course as original learning (OL) and the measure taken at the end of the retention interval as retention (R). Their loss function is the difference between the OL measure and the retention R measure divided by the OL measure: (OL – R)/OL. They assumed that their measures are ratio-scaled. However, there is a growing consensus that commonly used test scores are only ordinal measures, not ratio or even interval measures. Mathematical differences based on ordinal measures are often misleading. Furthermore, the point corresponding to "zero retention" is poorly defined in most of the studies Semb and Ellis used. Most of the studies employ multiple-choice or true-false assessment instruments, and only 15 of 62 used pretests. Therefore, a respondent with no knowledge of the material would be expected to get a positive R score (e.g., 50% correct on a true-false test). Thus, the zero point on the measurement scale does not correspond to zero retention, making the use of ratios invalid.

In summary, the interval and ratio operations used by Semb and Ellis (1994) with ordinal data make their findings difficult to interpret on the question of how much original learning is retained. Thus, although Bahrick (1984) and Semb and Ellis may offer reason for hope about the longevity of consumer behavior memories, the limitations of each of these studies raise the possibility that these memories are much shorter lived.

Additional macro-level evidence from the business literature supports the contention that memories of business education are indeed short-lived. Hunt, Chonko, and Wood (1986) found that students who majored in marketing were no more successful in marketing in the long term than those who majored in something else (using income, title, and several measures of job satisfaction as measures of success). Of course, there are other possible reasons why business education is not strongly linked to success in business careers, including the possibilities that marketing knowledge is not particularly useful in the workplace (Armstrong & Schultz, 1993), much of consumer behavior knowledge is little more than common sense (Armstrong, 1991), and many marketing pedagogies do not lead to substantial learning in the first place (cf. Chonko, 2004). In light of limitations of past research and the partial evidence that marketing education may not have lasting value, a rigorous exploration of the retention curve for consumer behavior knowledge is certainly warranted.

### The Shape of the Retention Curve

The general shape of the retention curve has been known for quite some time. The seminal work was conducted well over a century ago by Ebbinghaus (1885/1962; reviewed in Anderson, 2000), who determined that initially, forgetting is rapid, but then it continues at a decreasing rate (negative acceleration). Several more recent multisample studies have analyzed the shape of the retention curve using a much more comprehensive statistical approach and rigorous, laboratory-based experimental controls (e.g., Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). These studies generally find that a power function fits retention data as well or better than other similar functions (see also Gould, 2002). The retention curve generated by a power function (e.g., $y = x^k$, where $k < 0$ and $x$ is the retention interval) slopes steeply downward at the beginning and gradually becomes less steep, approaching a slope of 0 (a horizontal line) as $x$ increases. The power function will be used in this research to explore just how steep the retention curve is for one type of marketing-related learning.

In summary, the primary research question is:

*Research Question 1:* How fast do students forget what they learn in a consumer behavior course?

### The Process of Forgetting and Possible Remedies

Three main processes are believed to cause forgetting: decay, interference, and the absence of appropriate retrieval cues (Anderson, 2000; Bransford, 1979). The process of decay involves the spontaneous loss of a memory; knowledge once stored actually disappears entirely from memory. Although most researchers believe that decay plays a role in forgetting, many researchers assert that interference and the absence of appropriate retrieval cues are more important causes because retention can readily be manipulated using interference-related interventions, and memories that appear to be lost can actually be retrieved with the appropriate retrieval cues.

Interference refers to the interactions between two sets of memories. Proactive interference occurs when the knowledge of interest is affected negatively by a prior memory. For example, when a marketing student learns one framework for consumer decision making in introduction to marketing and another framework in consumer behavior, the first framework may interfere with the retention of the second. Retroactive interference occurs when the knowledge of interest is affected negatively by something learned subsequently, such as when two models of consumer motivation are taught in the same course. The student may not be able to distinguish the two well, and the second model may interfere with the memory of the first.

The third cause of forgetting is the absence of the appropriate retrieval cues. Many marketers will recognize this phenomenon in the difference between aided and unaided recall. With limited retrieval cues, the learner will be less likely to successfully access the knowledge of interest. For example, in Conway, Cohen, and Stanhope's (1991) research, former students were much more successful at recognizing the names of concepts they had once learned than they were at recalling (generating) names in an unaided context. In interpreting memory research results, it is imperative that the researcher pay close attention to the type of measure used because measures with more retrieval cues will generally yield higher retention scores.

The methods a teacher should employ to avoid interference and the lack of retrieval cues have been summarized well by several writers in the area. Ausubel (2000) advised teachers to help the learner establish a "stable trace" in memory. To be stable, the learner must be able to distinguish one concept from similar concepts so as to avoid interference. The learner must also connect the material to other relevant knowledge already possessed and thereby make the material meaningful and not simply rote memorization. The more connections the learner can make, the more anchored the trace will be in memory, thereby increasing the number of useful retrieval cues that will be available to the learner. As Neisser (1984) put it, "Information that is tied into an extensive and redundant cognitive structure . . . is sharply resistant to forgetting; isolated pieces of information, in contrast, are much more vulnerable" (p. 34). These observations about how to combat forgetting are reflected in the three pedagogical factors believed to improve retention that will be explored here: deep learning, repeated testing, and project-related learning.

### Deep Learning

Deep learning has gained attention in educational research ever since Marton and Säljö's (1976) seminal paper, although similar taxonomies of educational outcomes precede their work (e.g., Bloom, 1956; see Young, 2005, for recent work on this topic in marketing education). Marton and Säljö conceptualized deep learning as both a process and an outcome. Students who engage in deep learning strive to comprehend the material, generally engaging in elaborative processing to find additional meanings in the material. They go beyond just rote memorization (i.e., surface learning). They achieve a deeper understanding of the material and thus are more likely to score well on tests that measure comprehension, not just rote memorization. The surface-level outcomes and deeper outcomes described by Marton and Säljö are thus similar to the first two levels of Bloom's (1956) taxonomy, respectively. Bloom's first level, knowledge, is very similar to Marton and Säljö's surface learning (both models describe this level as "rote learning"). Bloom's next level, comprehension, is quite similar to Marton and Säljö's deep learning level (Marton and Säljö used the term *comprehending* to describe deep learning).

The Bloom (1956) model and the deep learning model differ on how rote learning and deeper learning are related. Bloom's taxonomy is a hierarchy, where higher-level knowledge depends on lower-level knowledge. If a student loses a lower-level element such as forgetting the definition of the lexicographic decision rule, he or she cannot identify which decision rule is being used in a particular situation. In Marton and Säljö's (1976) conceptualization, however, knowledge is not so hierarchical. An abstract form of the knowledge may be retained even if the details of the knowledge are no longer available. Neisser (1984) assumed a similar perspective in describing the schema theory of memory wherein a core concept may be retained in stable form even if details are lost. Following Bartlett's (1932/1967) schematic conceptualization of memory, Neisser suggested that we may not "recall" detail as much as we "recreate" detail based on our schemata and the available cues. This pattern of processing is one that marketing academics may have seen described as consumer inference (e.g., Dick, Chakravarti, & Biehal, 1990) or the systematic distortion hypothesis (Elliot & Roach, 1991). The schema theory of memory is consistent with findings from studies of the memory of stories, wherein learners may recall the roles of various characters but not their names (e.g., Stanhope, Cohen, & Conway, 1993). In the schema theory framework, a learner might forget some definitions but retain the ability to apply the concepts in a new situation.

Whether higher levels of knowledge are independent or dependent on lower levels of knowledge may depend somewhat on the specific knowledge in question. In their review of the very long-term memory literature, Semb and Ellis (1994) found that higher levels of understanding do not consistently affect retention. Such may be due to the many different types of knowledge contained in the studies reviewed. Some knowledge is conjunctive in nature such that the body of knowledge is only as useful as the least-remembered piece. In marketing research for example, students who forget how to interpret a $p$ value may lose the ability to form statistical conclusions even if the rest of their statistical knowledge is intact. Other knowledge is disjunctive such that the body of knowledge is useful if any one piece is recalled. For example, in consumer behavior, students who remember only one psychographic segmentation system (PRIZM or VALS) may be able to apply this knowledge to a situation to form managerial recommendations.

Additional variability may have been introduced across the Semb and Ellis (1994) studies related to the amount of elaboration of the higher-level knowledge. Student comments in Marton and Säljö (1976) indicate that deep learners look for meaning and connectedness of the material they study. This elaborative rehearsal may lead to a more extensive schema, or associative network, and thus a greater probability of recall. The deep learning process may also lead to schemata that are more distinct from similar schemata and therefore less prone to interference.

In summary, prior research results are somewhat inconsistent on the issue of levels of learning and retention. In the present context, the issue of retention of deep or surface learning may have more to do with the amount of elaboration involved. The amount of elaboration necessary to achieve higher-level consumer behavior learning—for example, finding additional examples, reworking homework exercises, and finding personal meanings—will likely lead to a stronger, more firmly anchored trace. Therefore,

> *Hypothesis 1:* Deep learning will be directly related to retention.

### Repeated Testing

Repeated exposures lead to greater learning (Halpern & Hakel, 2003), and when those exposures are spread over time they appear to be even more effective. Advertising researchers have been familiar with this phenomenon since the classic Zielske (1959) study that showed that spreading out advertising exposures over time leads to better recall. In an academic learning situation, Bahrick and Phelps (1987) showed that when Spanish vocabulary was learned over a longer time period (six to nine study sessions over 30 days vs. six to nine sessions back-to-back), the retention of this material 8 years later was increased. Conway et al. (1991) also hypothesized that their findings about the surprising stability of research methods retention were due to repeated learning—students learning and applying this material in several courses.

Students can be encouraged to spread out their study sessions by testing them several times on the same material via a series of cumulative exams throughout the term. In addition to spreading out the study sessions, this format may integrate a broader array of topics in the student's mind (Nungester & Duchastel, 1982), leading to more extensive schemata. In sum, the literature on retention consistently supports a relationship between repeated study over time and enhanced learning and retention. Therefore,

> *Hypothesis 2:* Retention will be directly related to repeated testing.

### Project-Related Learning

Several well-known education scholars have called for more active learning in higher education (Angelo, 1993; Chickering & Gamson, 1987; McKeachie, Pintrich, Lin, Smith, & Sharma, 1990), and several marketing education scholars agree (Hamer, 2000; L. K. Wright, Bitner, & Zeithaml, 1994). Examples of active learning pedagogies include projects, classroom exercises, and even classroom discussion, whereas the most common example of a passive learning pedagogy is lecture. Active learning enhances memory by encouraging the student to make additional connections with the target material. Rather than simply memorize definitions or record verbatim the content of a lecture, the student must often rearrange material and connect it to prior knowledge to complete the task at hand.

When projects are used as an active learning pedagogy, retention may increase for at least four reasons. First, students must identify how concepts from the text or lecture apply in another context. In so doing, they will make connections between the target content and the new context. Second, the project itself may then provide retrieval cues when students encounter a similar situation later. Third, projects often allow students to make choices about the direction of their learning experience, and these choices may enhance intrinsic motivation (Young, 2005). Finally, Conway et al. (1991) suggested that projects lead to the formation of procedural knowledge (i.e., how to do things), a fundamentally different form of knowledge than declarative knowledge (i.e., definitions and facts). They speculate that procedural knowledge itself may be more stable because it requires connecting several elements of declarative knowledge, and these connections make the traces more stable and accessible.

Because of the more elaborate mental connections that projects encourage, one would expect that content learning applied in a project is more likely to be retained than similar content that is not used in a project. Therefore,

> *Hypothesis 3:* Project-related learning will be directly related to retention.

## METHOD

The primary research question is addressed with the aggregate measures of learning and retention, and the follow-up hypotheses will be tested using subsets of the measures. Semb and Ellis's (1994) terminology is employed here for the key variables in this research, including original learning (OL), length of retention interval (RI), and retention (R), although their relative loss function is not used.

Rasch measurement (Rasch, 1960/1980), an advanced measurement approach that we have not seen used in any of the previous studies, is used here to assess OL and R. Rather than simply summing the number of correct responses on a test to compute a student's level of knowledge, Rasch measurement takes into account the differing level of difficulty of each question. By so doing, this approach offers two important advantages over previous studies. First, the knowledge measures have interval properties, and therefore curve fitting with these measures should lead to more valid results. Second, because estimates are generated of the difficulty of each test question, measures from tests that contain some but not all of the same items can be used to place students on the same latent measurement scale (i.e., test equating). A more thorough discussion of Rasch measurement is beyond the scope of this article, but interested readers will find Baker's (2001) book to be an excellent primer. A discussion of the potential for Rasch measurement in marketing can be found in Ewing, Salzberger, and Sinkovics (2005), and Bacon (2005) offered an example of its application in business education. For completeness, the technical results and fit statistics related to the Rasch model are included in the appendix. The data and analysis plan are described in more detail in the following.

### Data Collection

*Venue 1.* Original learning data were collected in a consumer behavior (CB) course over a 4-year period (11 class sections). The sections were taught by the same instructor (first author) at a moderate-sized private university in the Western United States. In total, 374 students completed a multiple-choice final exam in consumer behavior that was used as a measure of OL. Different versions of the final exam were used over this period, with the earliest exam comprising 67 questions and the most recent exam 107 questions. The latter versions did contain nearly all the questions from the earlier versions. This degree of overlap in questions across versions would be considered substantial in Rasch measurement, facilitating test equating. The instructor used a predominantly lecture style in all sections and used the Hawkins, Best, and Coney (2003) text. The sample comprised traditionally aged college students. Of these, 54% were women, 44% were marketing majors, 53% were third-year students, and 32% were fourth-year students. The mean grade measure among these 374 students was 85.16 (a straight B on a traditional scale), with a standard deviation of 10.62. The internal consistency reliability[1] of the OL measure was .89.

*Venue 2.* Retention data were collected in a marketing planning course (MP), which is the marketing capstone course (seven class sections) at the university. A subset of 50 of the questions from the CB final exam were extracted as a posttest instrument and administered to students who completed the test instrument in class for "extra credit." Students received no advance notice of the test. The data were collected over a 3-year period, corresponding approximately to the last 3 years of the collection of the CB final exam data. The internal consistency of the *R* measure was .74. Many of the students who completed the test instrument in MP had not taken CB at the research site, perhaps because they took an equivalent course at another school or because they were not marketing majors. Also, due to imperfect attendance, not all of the MP students completed the 50-question test. After matching the MP students to CB students, 92 complete sets of matched data were found. The demographics of this group were similar to the larger sample of 374 students who completed CB. The CB final grade measures were also similar to the larger sample, with a mean of 84.76 and a standard deviation of 9.34.

In addition to the 92 matched sets of data, 15 students were found to have completed MP and then later completed CB at the research site. The "retention" of this "control group" ($R_{CG}$) reflects the average level of knowledge that participants would have had at the end of the retention interval if they had never taken CB. (Other students who never took CB at the research site were not included in this control group because they may have taken CB at another school.) Thus, this $R_{CG}$ measure forms a baseline measure of knowledge. The measures from these 15 students were low ($M$ = 63.76, $SD$ = 4.73) but greater than chance (51.33). Thus, these students may have gleaned some knowledge of CB in college even without taking CB. In addition, perhaps some of what is taught in CB is common sense, and/or the better-than-chance scores demonstrate that some test questions provided extraneous cues that allowed clever students to guess with greater success than chance would suggest. When these students eventually took consumer behavior, the CB OL measures obtained ($M$ = 86.90, $SD$ = 11.94) were not substantially different from the matched sample, although 93% took CB as a senior.

The retention interval was computed for each student as the time between the completion of the CB final and the completion of the MP instrument. The CB final was always given in the 11th week (finals week) of the quarter, and the posttest in the capstone course was always given in the 7th week of the quarter. Thus, by examining the academic calendar, it was possible to estimate the RI in weeks. The RIs ranged from 8 to 101 weeks ($M$ = 39.92, $SD$ = 22.63, mode = 49) and included 12 unique intervals.

### Estimating the Retention Curve

The form of the power function of the retention curve used here is

$$R = OL(1 + RI)^k$$

where the value of k reflects the steepness of the retention curve over time. Within the parentheses, one unit is added to

RI so that the expected R will equal OL when RI equals zero. The exponent k can be estimated by taking the logarithm of each side of Equation 1 and applying linear regression.

**Testing Hypotheses About Differences in Retention**

Three pairs of subtests were identified in the final and in the posttest instrument that were related to each of the three hypotheses of interest. The sorting of test items into deep learning and surface learning categories was straightforward once a suitable sorting rule was established. Deep learning items included test questions that required students to identify the appropriate concept in a given scenario, identify the appropriate scenario given a concept, solve a problem using a concept, or recognize the similarities or differences between concepts. Surface learning items included test questions that asked about definitions or that could be mastered simply by rote memorization of the instructor's PowerPoint slides. The course instructor and one rater each independently categorized all the items as deep learning or surface learning. The instructor and the rater initially agreed on 44 out of 50 (88%) of the items (Cohen's [1960] kappa = .73; Landis and Koch, 1977, described this level of agreement as "substantial"). The 6 ambiguous items were then discussed, and agreement was reached on each.

The sorting of items into those that involved repeated testing and those that did not was also straightforward. The course syllabus provides a detailed list of learning objectives, including an exhaustive list of course concepts. Two or three cumulative exams were used in the course, depending on the quarter, and the scope of each exam was noted in the syllabus. Later exams focused more on the most recent material. Thus, on the final exam, some concepts were tested for the first time, whereas others were being tested for the second or third time. Those items being tested for the second or third time were considered to require repeated study, whereas those items being tested for the first time were considered to require study at one point in time. Of course, actual student study behavior may vary, but anecdotal conversations with students indicated that most students reviewed previous material when studying for the final exam.

Test items were categorized as either project related or not project related after close examination of the project descriptions and grading rubrics. Two projects were assigned in the course, so the project-related content included content from either project. The first project was completed individually by all students, and the second project was completed individually by a few students and in pairs by most students. Both projects required students to analyze a friend's decision-making process for a recent purchase. Several specific tools had to be applied (different tools on each project), and students were given the project scoring rubric in advance as part of the syllabus.

The subtest measures for deep learning, surface learning, material tested more than once, material tested once, project-related content, and non-project-related content exhibited adequate reliability given the statistical procedures used to test hypotheses. Subtest measures were computed for the CB final and posttest using the calibrations from the calibration sample described previously. The length of the subtests and their reliabilities are shown in Table 1. These reliabilities are generally lower than one would like (ranging from .43 to .66), but they are suitable for the early stages of research (Peter, 1979).

**TABLE 1**
**EXPLORATORY SUBTEST CHARACTERISTICS**

| *Subtest* | *Items* | *Reliability* |
|---|---|---|
| Deep | 17 | .43 |
| Surface | 33 | .66 |
| Tested more than once | 21 | .54 |
| Tested once | 29 | .64 |
| Project related | 22 | .58 |
| Not project related | 28 | .58 |

NOTE: Reliabilities were computed using the Rasch model but are analogous to Cronbach's alpha.

To test for differences in retention, the retention curve estimated to address the primary research question was used first to partial out the variance due to OL and RI. The residuals for each subtest computed from this model were then compared using paired $t$ tests to determine if there were significant differences in retention related to deep learning, repeated testing, or project-related learning. The use of the OL and RI measures essentially as covariates in these tests of differences in retention adds statistical power that substantially compensates for the modest reliability in the measures (Bacon, 2004).

## RESULTS AND DISCUSSION

The description of the results begins with an examination of the primary research question, which is essentially an exploration of the steepness of the retention curve for consumer behavior knowledge. Once this curve is established, deviations from this curve are evaluated related to deep learning, repeated testing, and project-related learning (Hypotheses 1, 2, and 3).

To estimate the retention curve of consumer behavior knowledge, the exponent in the power function shown in Equation 1 was first estimated by taking the logarithm of each side and applying regression analysis. The initial results were unexpected and prompted reconsideration of the model. The constant term estimate for Equation 1 was 20.6 and was statistically significant, $t(90) = 3.74$, $p \leq .001$, meaning the curve did not intersect the y-axis at R = OL when RI = 0. This finding lacks face validity because a retention interval of zero implies the exact moment the final exam was taken, and therefore the retention should equal the original learning. The functional form was then modified slightly to reflect the possibility that the shape of the retention curve is influenced by the amount of original learning. Thus, the new model to estimate was

$$R = OL(1 + RI)^{kOL}$$

The exponent k was again estimated by taking the logarithm of each side and applying regression. This model revealed a statistically significant exponent of $k = -.0058$; $t(89) = -4.85$, $p \leq .001$. Inserting the exponent back into Equation 2 and reestimating the regression indicated that the constant term was not statistically significant, $t(90) = 1.80$, $p = .075$, and the $R$-squared was .51. An inspection of the residuals indicated they were normally distributed and homoskedastic.[2] The exact form of this model was

$$R = 1.00OL(1 + RI)^{-.0058OL}$$

To better understand what this finding implies about the retention of consumer behavior knowledge, the expected retention curves for hypothetical A, B, and C students (final grade measures of 95, 85, and 75, respectively) are shown in Figure 1. The level of knowledge an average student would be expected to display if he or she had never taken consumer behavior (the baseline, or $R_{CG}$) is about 64 on the same scale. As can be seen in the figure, knowledge is lost very fast in the weeks immediately following the final. An average student (a final grade measure of 85) would be halfway (R measure of 74.5) to a control group score of 64 in just 13 weeks. The retention curve then appears to flatten somewhat, reflecting the negative acceleration of forgetting observed in so many previous studies. Still, after 2 years, the B students would be expected to score only a 68, just a few points above the baseline of 64. Interestingly, although the A students have higher achievement at the beginning, they lose knowledge at a faster rate than C students, and thus the difference between A student knowledge and C student knowledge is much smaller after 2 years.

The results are consistent with Hunt et al.'s (1986) findings about the long-term value of marketing education in general. However, they are more pessimistic about the longevity of consumer behavior memories than are findings of other research in the very long-term memory literature. Several reasons likely account for the lower retention rates reported here. First, although Bahrick's (1984) results imply that consumer behavior knowledge might last 40 years or more, a close read of Bahrick would indicate that if a student has only one course on the subject, the knowledge is not retained for more than a few years. Thus, the limited exposure that students typically have to consumer behavior knowledge may explain why the retention is relatively brief. Semb and Ellis (1994) were more optimistic about the retention of knowledge learned in school among the research they

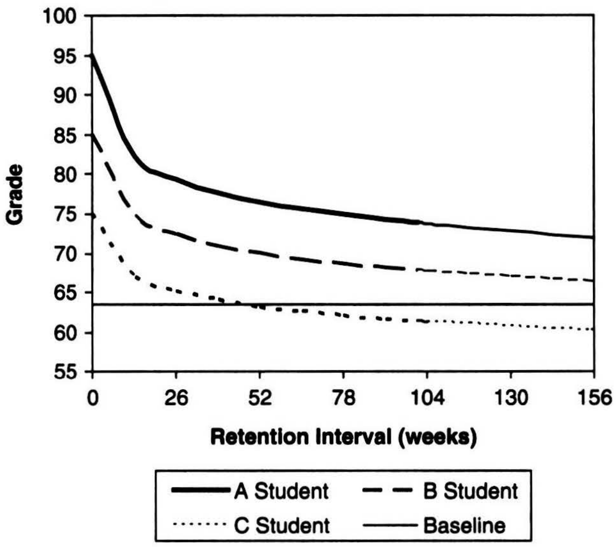


**FIGURE 1: Expected Grade by Retention Interval for Hypothetical Students**
NOTE: Lines are shown thinner after 101 weeks to reflect projections beyond the data in hand.

reviewed, but they used liberal assumptions about the quality of their measures.

**Tests of Hypotheses**

The three follow-up hypotheses about differences in retention were tested using the curve implied by Equation 3 as a baseline. The residuals from this line indicate weaker or stronger retention than expected, and so these residuals can be used to test hypotheses about factors that affect retention. Essentially, we control for differences in original learning and retention interval and then examine differences related to deep learning, repeated testing, and project-related learning. As shown in Table 2, two of the three hypotheses were supported. Hypothesis 1 was supported as knowledge learned at a deeper level was retained better than knowledge learned at a surface level. It must be noted however that the test items used were not nested within the same possible hierarchy. The surface learning items did not correspond to exactly the same concepts that the deep learning items captured, so one cannot conclude from these results alone that surface learning can be lost but deep learning retained within the same topic. Instead, the analysis suggests simply but importantly that once something is learned at a deeper level, it is more likely to be retained. The stability of these memories may be due in part to the elaborative processing necessary to achieve the deep learning in the first place. It is also important to note however that the effect size associated with deep learning (.35) was modest in Cohen's (1977) terms (.20 = small, .50 = medium). Even when disattenuated for measurement error, the effect size does not exceed what

TABLE 2
COMPARISON OF RESIDUALS ACROSS SUBTESTS

| *Subtest* | M | *Difference* | *Pooled* SD | *Effect Size* | p |
|---|---|---|---|---|---|
| Deep | 2.43 | 3.24 | 9.30 | 0.35 | .01 |
| Surface | –0.81 | | | | |
| Tested more than once | 2.30 | 3.64 | 9.41 | 0.39 | .01 |
| Tested once | –1.34 | | | | |
| Project related | 0.43 | 0.54 | 8.58 | 0.06 | .66 |
| Not project related | –0.10 | | | | |

NOTE: The *p* values reflect the results of two-tailed paired *t* tests. Difference scores are computed on a traditional grading scale (e.g., 90 to 100 = A, 80 to 89 = B, etc.). Thus, a difference of 3.24 represents about a third of a letter grade.

Cohen would describe as medium. With the traditional grading scale used here, where 10 points out of 100 reflects one full letter grade, the difference in grades of 3.24 amounts to approximately a third of a grade.

Hypothesis 2, relating to repeated testing, was also supported. This effect (.39) was about the same size as the effect associated with deep learning. This particular finding may help explain why initial learning was associated with the slope of the retention curve, whereas in many other studies, retention curves are approximately parallel. As Young (2005) demonstrated, students learn in a variety of ways. All students learn some material in a series of "study sessions over time" as they listen to material covered in class. Most students probably also learn material over a short period of time when they study for the exam. It is possible and consistent with the first author's anecdotal observations that students who scored the highest on the final exam did so by "cramming" heavily shortly before the exam. Young's results also suggest that extrinsically motivated students, which may translate to grade-oriented students, learn a proportionately larger amount of whatever they learn using rote memorization, perhaps over a short period of time. This knowledge is more vulnerable to various forgetting processes than the material studied in more complex ways over a longer period of time. Consequently, the higher-scoring students have steeper retention curves, but because they have also learned in some of the same ways as the weaker students have learned, the retention among the stronger students never drops below the retention of the weaker students. This pattern of nonparallel retention curves may not have emerged in previous research because so much of the research on memory is conducted in laboratory settings, where experimental participants learn in ways that are carefully controlled.

Hypothesis 3, relating to project-related learning, was not supported. Although the effect was in the expected direction, the effect was small and did not achieve statistical significance. The observed effect size may have been even smaller if the subscales used (deep/surface, one test/repeated testing, and project/nonproject) were completely independent. However, many of the deep learning items and the repeated testing items were also project-related learning items. The item membership correlation between deep learning and project-related learning was .38, and the correlation between repeated testing items and project-related learning items was .47. Thus, any positive findings related to the project-related learning might also be because several project questions were also deep learning questions or repeated testing questions. The item membership correlation between the deep learning items and the repeated testing items was somewhat lower, at .24.

The null results may be due to the variability of the coverage of the project-related content across projects and students. For example, although the project required students to identify and describe how a decision rule was used, the exact decision rule identified varied across projects. Thus, some students may learn the conjunctive rule in greater depth, whereas other students learn the lexicographic rule in greater depth. On the exam, any question related to the conjunctive or lexicographic rule was considered to be project related in addition to any question related to the disjunctive, elimination-by-aspects, or compensatory rule or any question that asked about a combination of these rules. Thus, the test items that were categorized as project related may not have been related to all of the students' projects. There were other sets of concepts on the project (e.g., modes of problem recognition, type of reference group influence, methods of information search, etc.) that followed this pattern. Therefore, although the present results cannot be interpreted as supporting the contention that projects lead to greater retention, the analysis does not strongly deny the possibility that projects are associated with retention. However, if one were to accept that projects were not associated with retention here because different students learned different things on different projects, then one must also accept that projects represent a kind of hit-or-miss method of learning, which may be troublesome in itself.

## IMPLICATIONS AND RECOMMENDATIONS

Based on the study's findings, the following recommendations are offered.

*Develop a pedagogy that requires deep learning early and often.* The results reported here demonstrate that material learned at a deeper level will be retained longer than material learned at a surface level. To increase the long-term value of a student's education, it follows that techniques that promote deeper learning and more elaborative processing should be used more often. For example, Michaelsen's "team learning" approach can be used when students are tested on material for the first time (Michaelsen & Black, 1994). The group quiz component of Michaelsen's method promotes a student discussion of the quiz questions that can lead to deeper learning. This technique could then be followed with semistructured classroom-based experiential learning techniques (Hamer, 2000) to help students gain greater insight into the material (other techniques are referenced in Young, 2005). Students may also need to be taught how to study in ways that encourage elaborative processing and achieve deep learning. For example, students could be taught to use the PQ4R approach to deep studying, which involves previewing material, identifying questions, reading once, reflecting, recalling out loud, and reviewing as necessary (Hartley, 2002, discussed this and other study skills worth teaching marketing students). Ideally, this sequence of active learning exercises would continue beyond any given course and into several subsequent courses, offering students the repetition, cross-course integration, and consequently increasing depth necessary to form complex knowledge structures that they will retain for years to come.

*Sacrifice breadth for depth.* Researchers have long known that pedagogies that promote deeper learning may require more time (Hilgard, Irvine, & Whipple, 1953). To allow more time for the elaboration and depth necessary to achieve increased retention, course designers will need to give up some breadth in topics covered. Marketing faculty should engage in longer conversations with each other, perhaps in the context of assessment and learning goal setting, to agree on which topics are the most important and which topics are dispensable. It is important to remember that although we hate to "give up" some of our favorite topics, the topics that are only covered in passing are not meaningfully retained. Thus, we have already been giving them up; it just has not been obvious. To avoid giving up everything, a few important topics must be covered more in depth.

*Require that students take a course's prerequisites immediately before the course.* Curricula should be carefully reevaluated in light of the evidence that learning may only be retained for a short period of time. Basic courses often teach concepts, and later courses apply the concepts. However, unless the follow-up course draws on the initial knowledge within a few months of initial exposure, most of the initial learning will be lost. For example, when students take introduction to statistics 2 years before marketing research, the students will have forgotten most of their statistics by the time they need it, and so the prerequisite is of little value. Rather than simply requiring prerequisites, course designers should consider requiring that certain prerequisites be taken immediately prior to a course.

*Focus course content on concepts and tools that students will encounter in their first job.* Whereas the results presented here indicate that knowledge is quickly lost, such knowledge might be recalled or relearned quickly in the workplace if the appropriate retrieval cues are present. This possibility implies that we should only teach tools and concepts that students will likely encounter in their first job. Knowledge that is not quickly refreshed in the workplace will soon be forgotten and consequently lost forever. In some cases, the instructor may choose to teach tools that are not commonly used. In these cases, the students must be able to learn the tools at such a deep level and perhaps with repeated applications that the memory will be stable and accessible even without strong external cues. Such learning would take a substantial investment in resources and so should not be taken lightly.

*Use cumulative exams.* The findings strongly support the use of cumulative tests when tests are used. In an ideal program, tests or other assessments would even be cumulative across courses to provide impetus for students to relearn and reintegrate material as they progress through the program. Taken one step further, a series of courses would use the same materials (e.g., same texts or text) to facilitate cross-course cumulative assessments and to eliminate the memory interference that can occur when students learn similar but slightly different content in different courses.

*Assume a broader approach to teaching evaluation.* Unfortunately, cumulative exams are likely to be unpopular with students. The first author has received several comments on student evaluation forms that complain about the cumulative final exam. Thus, the use of cumulative exams represents a classic conflict between an instructor's need to secure high teaching evaluations and the desire for students to achieve long-lasting learning. This conflict underscores the need for a broader approach to teaching evaluation, perhaps including teaching portfolios (Babin, Shaffer, & Tomas, 2002), so that teachers will not be punished for using methods that actually increase long-term retention.

*Implement learning assessments across a wider time frame.* The study findings imply that most assessment measures fail to capture what matters most: very long-term retention. Most course-embedded, grade-related assessments will only capture what is known on the last day of class. If we are to assess what matters most, we must assess across a wider time frame. Furthermore, the assessment system should include comprehensive assessments that are not graded or preannounced (like the retention test used here). Grading and preannouncement of assessments would lead to cramming, which would give artificially high estimates of very long-term retention, undermining the purpose of the assessment.

### Limitations, Recommendations for Future Research, and Conclusions

Several aspects of the research design used here should be carefully considered in interpreting the results. First, this study used what many consider to be recall measures. For example, some questions required students to apply the conjunctive rule in a new context. Other studies have used recognition measures, where the instrument only asked students if they had ever heard the term *conjunctive rule* before. Students will generally score higher on recognition measures than on recall measures (see Conway et al., 1991), indicating that some trace of knowledge still exists in memory even when the trace is not sufficient to correctly answer some types of questions. Additional evidence that memorial traces exist after learning appears to be forgotten comes from the observation, first reported by Ebbinghaus (1885/1964), that knowledge that appears to be lost can be relearned more quickly than new knowledge can (see also Nelson, 1971, for a more recent example). Thus, in an environment with the appropriate cues and opportunities, students may recall or relearn information and thus perform at a higher level than the results presented here would suggest.

Two aspects of the measurement of learning in this research likely have countervailing effects on the results. First, the OL measure and the R measure both used the same questions. Thus, although none of the students were informed of the correct answers, some students may have recalled some of the questions and even their deliberations at the time of the final. The possibility of "remembering to the test" may have inflated the R scores. In side conversations, several students admitted they remembered some questions, but none admitted to remembering the exact answer they previously selected (cf. Nungester & Duchastel, 1982). A factor that may have deflated the R scores was that there was not a performance reward for the posttest. The disparity between the graded nature of the final and the volunteer nature of the posttest likely led to differing levels of effort and therefore differing levels of performance on the test. The extent to which remembering to the test may balance the lack of a performance reward is unclear and presents an opportunity for future research.

Another limitation of this research is that all sections of the course were taught by one instructor with one instructional style (lecture with some discussion). Although this is a common method of teaching in marketing (Clow & Wachter, 1996; Roach, Johnston, & Hair, 1993; Smart, Kelly, & Conant, 2003), other styles may lead to different retention curves. Evidence is accumulating that group learning exercises (Johnson & Johnson, 2002) or exercises that motivate deep learning (Marton & Säljö, 1976; see also Hamer, 2000; L. K. Wright et al., 1994) may lead to greater long-term retention. Although substantial research exists in marketing on how students feel about various teaching styles (e.g., Smart et al., 2003), the effect of teaching style or the application of study skills on very long-term retention is an important area for future research.

One unexpected finding related to teaching was the lack of significant differences in retention related to project learning. As stated previously, project learning was expected to lead to greater elaboration and perhaps the acquisition of procedural knowledge, thus leading to greater retention. The measures used here however did not capture the kinds of idiosyncratic elaborations that might occur in projects and did not strongly tap procedural knowledge. Future research on project effectiveness should consider alternative measures of learning to capture these outcomes.

Another important issue to consider for future research is the conceptualization measurement of learning outcomes themselves. Whereas multiple-choice tests have been shown to correlate highly with other forms of testing such as short answer or essay tests (Bacon, 2003; Lukhele, Thissen, & Wainer, 1994), the possibility remains that students are somehow changed in a more fundamental way as they progress through a marketing program. Perhaps projects and other experiences help students develop critical thinking skills, team skills, business perspective, confidence, or a sense of self as a professional. The effects of a marketing curriculum on these learning outcomes is an important area for future research.

In conclusion, the research presented here suggests that most consumer behavior knowledge is forgotten soon after a student completes the course. This research also found that knowledge learned at a deeper level and knowledge subjected to repeated testing were retained longer than other knowledge. By understanding how much learning is lost, marketing educators can begin to find ways to minimize the loss and thus maximize the long-term value of the education they provide.

## APPENDIX
## Rasch Measurement Results

Rasch measurement was used in the present research to create measures of original learning and retention among the study participants. To apply Rasch measurement, the difficulty of the test questions (items) on the final was first calibrated using a large sample, and these calibrations were then applied to the observations from the smaller matched sample to form measures of retention. Rasch software reports these measures in units called *logits*. It is common in Rasch research to then rescale these logit measures to a more meaningful scale by applying a linear transformation. The details of the calibration and rescaling are included in this appendix.

The sample of 374 students who had completed the consumer behavior (CB) final exam was used to calibrate the final exam. WINSTEPS software was used in all Rasch analyses presented here (Linacre, 2004). In the calibration process, the exact difficulty of each test question (item) is estimated simultaneously with the original learning (OL) level of each student. The item calibrations can then be applied to the posttest instrument sample, allowing the estimation of the retention (R) measure of each of the market planning course (MP) students. By using the same calibrations, all of the OL and R measures will be on the same latent scale. Because of the substantial overlap among the versions of the final, the procedure known as the "simple procedure" (Bond & Fox, 2001, p. 54) was used for test equating.

Fit statistics from the calibration sample confirmed the appropriateness of the Rasch model in this application. Student information weighted fit statistics ranged from 0.76 to 1.39. B. D. Wright, Linacre, Gustafson, and Martin-Löf (1994) suggested that a range of 0.7 to 1.3 or less is appropriate for a run-of-the-mill (not high-stakes) multiple-choice test; item fit statistics had a slightly better fit.

Rasch measurements are reported in units of logits, which are then commonly transformed into units that may be more meaningful to the researcher. The student OL measures (i.e., student final exam measures) were distributed with $M = .948$ and $SD = .828$ logits. For ease of interpretation, these Rasch measures[3] were converted back to approximate 100-point-scale grades using a linear transformation. Whereas Rasch measures are a nonlinear function of classical test theory (CTT) scores, the correlations among Rasch measures and CTT scores are generally quite high, and so the conversion is a fairly close approximation to the CTT-based grades that were actually given on the exams. The Rasch grades measures in the calibration sample were distributed with $M = 85.16$ and $SD = 10.62$. The internal consistency reliability of the OL measure was .89.

## MEASURING RETENTION

The item difficulties estimated from the calibration sample were next used to measure R in the posttest sample. The student fit statistics for these R measures were similar to the fit statistics in the calibration (OL) sample (range = 0.67 to 1.35), but the item fits were worse (range = 0.54 to 3.29), suggesting differential item functioning (DIF). Such DIF would be expected if some types of material were forgotten faster than other types of material. The internal consistency of the R measure was .74. The students in the matched sample achieved scores on the OL measure that were similar to the calibration sample (final grade measure $M = 84.76$, $SD = 9.34$).

## NOTES

1. All of the internal consistency reliabilities reported here are generated from the WINSTEPS program following the Rasch model (Linacre, 2004). These coefficients are analogous to Cronbach's alpha, although the specific formula used is different due to the use of the Rasch model.

2. The difference in residuals associated with gender was not found to be statistically significant, mean difference = .159, $t(90) = .126$, $p = .90$, implying retention was not directly related to gender. Thus, although previous studies have found differences in original learning related to gender (e.g., Bacon, 2003), after controlling for differences in original learning, the retention curves of men and women appear to be similar.

3. Following the tradition in the Rasch literature, the term *measure* is used throughout to refer to intervally scaled metrics derived from Rasch measurement. The term *score* is used to refer to classical test theory metrics, which may only have ordinal properties.

## REFERENCES


Anderson, J. R. (2000). *Learning and memory* (2nd ed.). New York: John Wiley.

Angelo, T. A. (1993). A "teacher's dozen": Fourteen general, research-based principles for improved higher learning in our classrooms. *AAHE Bulletin, 45*(8), 3-13.

Armstrong, J. S. (1991). Prediction of consumer behavior by experts and novices. *Journal of Consumer Research, 18*, 251-256.

Armstrong, J. S., & Schultz, R. L. (1993). Principles involving marketing policies: An empirical assessment. *Marketing Letters, 4*, 253-265.

Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Norwell, MA: Kluwer Academic Publishers.

Babin, L. A., Shaffer, T. R., & Tomas, A. M. (2002). Teaching portfolios: Uses and development. *Journal of Marketing Education, 24*, 35-42.

Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education, 25*, 31-36.

Bacon, D. R. (2004). The contributions of reliability and pretests to effective evaluation. *Practical Assessment, Research & Evaluation, 9*(3). Retrieved April 3, 2004, from http://PAREonline.net/getvn.asp?v=9&n=3

Bacon, D. R. (2005). The effect of group projects on content-related learning outcomes. *Journal of Management Education, 29*, 248-267.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology-General, 113*, 1-31.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54-75.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology–Learning, Memory and Cognition, 13*, 344-349.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Bartlett, F. C. (1967). *Remembering*. Cambridge, UK: Cambridge University Press.

Bloom, B. S. (Ed.). (1956) *Taxonomy of educational objectives, Handbook I: Cognitive domain*. New York: Longman, Green and Company.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Bransford, J. D. (1979). *Human cognition: Learning, understanding, and remembering*. Belmont, CA: Wadsworth.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association of Higher Education Bulletin, 39*(7), 3-7.

Chonko, L. B. (2004). If it walks like a duck . . . : Concerns about quackery in marketing education. *Journal of Marketing Education, 26*, 4-16.

Clow, K. E., & Wachter, M. K. (1996). Teaching methodologies used in basic marketing: An empirical investigation. *Journal of Marketing Education, 18*, 48-59.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). San Diego: Academic Press.

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–General, 120*, 395-409.

Dick, A., Chakravarti, D., & Biehal, G. (1990). Memory-based inferences during consumer choice. *Journal of Consumer Research, 17*(1), 82-93.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Original work published 1885)

Elliott, K. M., & Roach, D. W. (1991). Are consumers evaluating your products the way you think and hope they are? *Journal of Consumer Marketing, 8*(2), 5-14.

Ewing, M. T., Salzberger, T., & Sinkovics, R. R. (2005). An alternative approach to assessing cross-cultural measurement equivalence in advertising research. *Journal of Advertising*, *34*(1), 17-36.

Gould, S. T. (2002). Simple power patterns in response to advertising. *Journal of Advertising Research*, *42*(6), 55-68.

Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning to the university and beyond. *Change*, *35*(4), 36-41.

Hamer, L. O. (2000). The additive effects of semistuctured classroom activities on student learning: An application of classroom-based experiential learning techniques. *Journal of Marketing Education*, *22*, 25-34.

Hartley, J. (2002). Studying for the future. *Journal of Further and Higher Education*, *26*, 207-227.

Hawkins, D. I., Best, R. J., & Coney, K. A. (2003). *Consumer behavior: Building marketing strategy* (9th ed.). Boston: Irwin/McGraw-Hill.

Hilgard, E. R., Irvine, R. P., & Whipple, J. E. (1953). Rote memorization, understanding, and transfer: An extension of Katona's card-trick experiments. *Journal of Experimental Psychology*, *46*, 288-292.

Hunt, S. D., Chonko, L. B., & Wood, V. R. (1986). Marketing education and marketing success: Are they related? *Journal of Marketing Education*, *8*(2), 2-13.

Johnson, D. W., & Johnson, R. T. (2002). Social interdependence theory and university instruction: Theory into practice. *Swiss Journal of Psychology*, *61*, 119-129.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Linacre, J. M. (2004). *A user's guide to WINSTEPS: Rasch-model computer programs*. Chicago: MESA Press.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, *31*, 234-250.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I-Outcome and process. *British Journal of Educational Psychology*, *46*, 4-11.

McKeachie, W. J., Pintrich, P. R., Lin, Y-G., Smith, D. A. F., & Sharma, R. (1990). *Teaching and learning in the college classroom: A review of the research literature* (2nd ed.). Ann Arbor, MI: National Center for Research to Improve Post Secondary Teaching and Learning.

Michaelsen, L. K., & Black, R. H. (1994). Building learning teams: The key to harnessing the power of small groups in higher education. In S. Kadel & J. Keehner (Eds.), *Collaborative learning: A sourcebook for higher education* (Vol. 2, pp. 65-81). State College, PA: National Center for Teaching Learning and Assessment.

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology–General*, *113*, 32-35.

Nelson, T. O. (1971). Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 568-576.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*, 18-22.

Peter, J. P. (1979, February). Reliability: A review of psychometric basics and recent marketing practices. *Journal of Marketing Research*, *16*, 6-17.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press. (Original work published 1960).

Roach, S. S., Johnston, M. W., & Hair, J. F., Jr. (1993). An exploratory examination of teaching styles currently employed in marketing education: Developing a typology and its implications for marketing students. *Journal of Marketing Education*, *15*(3), 32-38.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734-760.

Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research*, *64*, 253-286.

Smart, D. T., Kelly, C. A., & Conant, J. S. (2003). Mastering the art of teaching: Pursuing excellence in a new millennium. *Journal of Marketing Education*, *25*, 71-78.

Stanhope, N., Cohen, G., & Conway, M. (1993). Very long-term retention of a novel. *Applied Cognitive Psychology*, *7*, 239-256.

Wixted, J. T., & Ebbesen, E. G. (1991). On the form of forgetting. *Psychological Science*, 2, 409-415.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*, 370.

Wright, L. K., Bitner, M. J., & Zeithaml, V. A. (1994). Paradigm shifts in business education: Using active learning to deliver services marketing content. *Journal of Marketing Education*, *16*(3), 5-19.

Young, M. (2005). The motivational effects of the classroom environment in facilitating self-regulated learning. *Journal of Marketing Education*, *27*, 25-40.

Zielske, H. A. (1959). The remembering and forgetting of advertising. *Journal of Marketing*, *23*, 239-243.