

Effects of Test Taking on Retention Learning in Technology Education: A Meta-Analysis

W. J. Haynie, III

Commencing in 1985, a small body of experimental studies on the effects of test taking on delayed retention learning of technical subject matter has been completed in technology education settings. Much of the learning in technology education courses, especially the hands-on aspects, are best assessed via instruments and techniques other than traditional tests; but classroom tests are still important for learning in the cognitive domain and the time that technology teachers spend administering tests is best spent if the tests also help students learn.

Two of the studies were completed in public schools and the others were conducted in university classes. One central question in all of the studies concerned the effects of taking tests on the delayed retention of the information tested. Delayed retention is important because that comprises the information and concepts that the student still knows three or more weeks after the effects of “cramming” for the test have evaporated—thus, delayed retention represents the important and significant learning in a technology course. Nine other related factors were examined concerning the types of tests used, time on task, and subtle variations in the setting of the experiments. A total of nine experiments were conducted over a twenty year period. One of the studies failed and the other eight had significant findings of importance in answering the research questions posed. This article reports a meta-analysis conducted by the author of the series of studies to summarize that body of work.

Background

Testing in education, including technology education, and the large amounts of instructional time it consumes are important topics of continuing research. Most of the research reported on testing has historically concerned standardized tests (Stiggins, Conklin, and Bridgeford, 1986; Haynie, 2004). Today, with high stakes testing employed to track student and school performance, this emphasis on standardized testing in the research literature continues. Most of the evaluation done in schools, however, is done with teacher-made tests (Haynie, 1983, 1990a, 1992, 1994, 1997b, 2003a, & 2004;

W. J. Haynie, III is Professor in the Department of Mathematics, Science, and Technology Education and Coordinator of Technology Education at North Carolina State University, Raleigh.

Herman & Dorr-Bremme, 1982; Mehrens, 1987; Mehrens & Lehmann, 1987; Newman & Stallings, 1982). Research efforts on the effects of teacher-made tests and other issues surrounding them such as frequency of use, quality, investment of the time required to administer and return them, benefits for student learning, optimal types to employ, and usefulness in evaluation add valuable findings to the body of knowledge for educators. The available findings on the quality of teacher-made tests cast some doubt on the ability of teachers to perform evaluation effectively (Burdin, 1982; Carter, 1984; Fleming & Chambers, 1983; Gullickson & Ellwein, 1985; Haynie, 1983, 1992, 1997b; Stiggins & Bridgeford, 1985). Mehrens and Lehmann (1987) cite the importance of teacher-made tests in the classroom and their special ability to be tailored to specific instructional objectives. Evaluation through teacher-made tests in schools is an important and needed part of the educational system (including technology education classes) and a crucial area for research (Haynie, 1983, 1990a, 1992, 2003a; Mehrens & Lehmann, 1987).

Marsh (1984) reported that, despite fears of tests and distaste for them, students in an experimental study self-reported that they studied more for material which was to be tested in-class. This finding was consistent across groups—both those who had been tested in-class and those who had been tested by take-home methods reported that they learn more when facing an in-class test. Mixed findings on the anxiety caused by tests have been reported by Denny, Paterson, and Feldhusen (1964), and Marsh (1984). Several studies have demonstrated the positive effects of in-class tests on retention (Duchastel, 1981; Haynie, 1990b, 1991, 1994, 1995, 1997a, 2003a, 2004; Nungester & Duchastel, 1982). All of the studies by Haynie were completed in technology education settings with technical subject matter. Some studies have shown positive effects of reviews instead of tests on retention (Haynie, 1990a; and Nungester & Duchastel, 1982). Haynie (1995) showed some benefits of post-test reviews on retention.

Haynie (1990a) initiated a series of experimental studies to examine the issues concerning teacher-made tests and their effects on retention learning in technology education. A protocol was developed following the general methodology of Nungester & Duchastel (1982) with only those changes needed to adapt it to a technology education setting. The subject matter consisted of concepts, factual information, and applications concerning modern “high tech” materials (including composites, exotic metals, and heat shielding materials) developed or used in the NASA space exploration program. The information was presented via videotapes, including live demonstrations, script written by the researcher, excerpts from NASA “Tech Briefs” and other publications, and graphs/charts showing the relationships of various characteristics of the materials under study. Due to two unrelated factors, this first study failed to make any significant findings on the research questions posed concerning the learning effects of testing. One problem was a weakness of the 20-item test instrument used to collect the data and the other was failure of teachers to follow important directions crucial to the design of the study. The teachers were

instructed to isolate participants from distracting activities nearby in their regular open-space lab by using another room pre-arranged by the researcher. Instead, half of the teachers showed the videotapes and administered the tests in the open-space lab while other groups worked on project construction activities—thus greatly confounding the experiment. Though there were some non-significant trends in the data from the teachers who had cooperated, overall there were no significant findings about the research questions posed. This study had some unrelated serendipitous findings concerning difficulties in presenting effective instruction in open-space environments which were reported in a minor journal (Haynie, 1990a). This study is not included in the meta-analysis presented here. Its importance, however, is that it laid the groundwork for the later successful investigations. The 30-item version of the instrument that resulted was used for the entire series of studies that followed. Eight successful experiments were then conducted and reported (Haynie 1990b, 1991, 1994, 1995, 1997a, 2003a, 2003b, and 2004). This meta-analysis examines the key findings of those eight studies.

Methodology of the Experiments

The protocol for the experiments, as approved by the university human subjects review panels (first at George Mason University and then at North Carolina State University) involved the initial instruction of all groups, a test or no-test treatment, a three-week delay period, and a final unannounced delayed retention test on the same information. Initial instruction in the first two efforts (conducted in public schools) was via a videotape developed by the researcher and his graduate students. All of the other studies were conducted in university classes with instruction via printed text materials. In the text-based studies, the information on “high tech” materials and their applications was presented in a booklet developed by the researcher that included original text, excerpts from NASA “Tech Briefs” and other publications, graphs/charts, and discussion of civilian applications of the materials studied. The booklet included a table of contents, text, halftone photographs, line drawings, and a full index to make it parallel with any regular course textbook. In the earliest studies all participants were informed that their scores on initial tests (if given) would not count in their course grade because the unit was newly added to the course and final materials were still in development (Haynie, 1990b, 1991, 1994, and 1995). This factor was criticized by reviewers of the early studies who questioned whether students had taken the unit of study seriously and had actually “given a good, honest effort” as requested during the directions for the unit. Therefore, some of the same questions were revisited in later studies (Haynie, 1997a, 2003 a, 2003b, and 2004) in which all students knew from the start that their grades would be affected by any tests taken in the unit (except, of course, the unannounced delayed retention test which was still voluntary and used for research purposes only).

The delayed retention test had 30 items. Twenty of these items were alternate forms of the items used in initial tests for those groups who were

initially tested. The remaining ten items comprised a subtest of novel information used to determine if students studied the entire booklet or simply hunted for the answers via the index (in cases where take-home tests were used or study questions were provided). Those ten items were interspersed so that students would not perceive them differently from the original twenty. The tests operated chiefly at levels 1-3 of Bloom's taxonomy: Recall of facts, conceptual understanding, and application of learning to novel situations. Each level was represented equally. The delayed retention test scores were the only data analyzed to answer the research questions.

Another factor criticized in some of the early studies concerned assurance of equality in ability of the groups participating. In the earliest studies (Haynie 1990a, 1990b, 1991, 1994, and 1995) random assignment was the only technique used to assure equal ability in the groups. Each experimental group was comprised of several intact class sections combined together to form one group. This technique provided an adequate n for the experiments to have acceptable power and also reduced the likelihood of extraneous variables such as time of day, semester, teacher (graduate assistant) conducting the class, or other factors from systematically affecting any particular experimental or control group. Since the course sections generally enrolled 20 or fewer students, two to four independent sections (randomly assigned) were required to make each experimental or control group (ranging from 35 to 71 depending upon how many groups were compared in each study). In all studies the n for the groups compared within a particular experiment was very similar.

The researcher felt that this randomization process was adequate assurance of equal entering ability among groups. However, following the advice of the most critical reviewers, in the final four studies (Haynie, 1997a, 2003a, 2003b, and 2004) the researcher demonstrated equality via a related "metals pretest" administered immediately before the experiment began. The study topic in the experiment involved high-tech and composite materials, so the metals pretest (covering the unit studied just prior to the experiment) was viewed as an adequate indicator of equal ability. In none of those experiments was a difference found in entering ability for either the experimental or control groups.

Normal precautions were taken to assure adequate lighting, temperature control, quiet atmosphere, limited distractions, and other comfort and privacy factors to provide an acceptable test environment. All directions concerning participation in the studies were read from prepared scripts to avoid confounding factors. The delayed retention test was carefully prepared and evaluated with reliability ratings between .69 and .74 for various studies within the series using Cronbach's Coefficient Alpha. According to Thorndike and Hagen (1977), tests with reliability approaching .70 are within the range of usefulness for studies of this type. All study materials were collected following the initial two-week instruction period and were maintained in secure storage to prevent advance information for future groups of students. In a debriefing session following the experiment, students were requested not to share any

information about the experiment, its methods or purposes, or the unit of study with their peers.

The factors involving the types of test or no-test conditions, use of study questions or reviews, and exactly what was announced prior to or during study of the unit of information were the various treatments in the investigations (independent variables) and the performance on the delayed retention test was the common dependent variable in all of the experiments. This consistency allows reasonable comparison of results in this meta-analysis. Readers who desire more specific details about the treatments used in any particular study are encouraged to read the original reports as published.

Methodology of the Meta-Analysis

The methodology of this meta-analysis involved calculation of the “Effect size” (Δ) for each factor of interest in all studies which examined that factor (Borg & Gall, 1989). The effect size was found using the formula:

$$\Delta = \frac{\bar{X}_{\text{Exp}} - \bar{X}_{\text{Con}}}{SD_{\text{Con}}}$$

Once the effect sizes were determined, they were also averaged with consideration of the n of each mean to find a weighted “Mean effect size”. The Effect sizes, Mean effect sizes, and Number of positive findings are reported for ten research questions of interest in Table 1, following procedures used by Mayer and Moreno (2002) in a similar effort. The remainder of this report examines the composite findings on these research questions.

Findings

Ten questions were considered in the eight experimental studies. All of the studies sought answers to two of those (generalized) questions:

- Does taking a test increase retention learning? (Factor 2 on Table 1), and
- Does time on task (including tests) increase retention learning? (Factor 3).

The remaining 8 factors (1, 4, 5, 6, 7, 8, 9, and 10) were only considered in one or two studies each. There were significant positive findings on six of the eight remaining related research questions with only factors 1 and 5 having no supportive significant findings. There were slight non-significant positive trends supporting factors 1 and 5. Discussion of each of the research questions and the studies which examined them follows.

Factor 1: Does prior knowledge of an upcoming test increase retention learning?

In the two experiments in which a group was told to expect a test but they did not actually receive a test, they showed only slightly higher achievement on delayed retention vs. groups told they would not be tested (Haynie, 1990b,

Table 1
Effects of Tests on Retention Learning: Meta-analysis of Eight Experimental Studies.

	Findings & Sources	Effect size Δ	Mean effect size	Number of positive findings[†]
1	Prior knowledge of upcoming tests increases retention learning		0.05	2 of 2
	1990b	0.01 NS		
	1997a	0.09 NS		
2	Taking a test increases retention learning		0.76	8 of 8
	1990b	0.54 *		
	1991	1.10 *		
	1994	0.90 *		
	1995	1.29 *		
	1997a	1.57 *		
	2003a	0.67 *		
	2003b	0.17 NS		
	2004	0.26 NS		
3	Time on task (including tests) increases retention learning		0.85	8 of 8
	1990b	0.60 *		
	1991	1.10 *		
	1994	0.90 *		
	1995	1.29 *		
	1997a	1.57 *		
	2003a	0.67 *		
	2003b	0.40 *		
	2004	0.26 NS		
4	Take-home tests support retention learning		0.58	2 of 2
	1991	1.08 *		
	2003b	0.08 NS		
5	Take-home tests support retention better than in-class tests		0.10	2 of 2
	1991	0.05 NS		
	2003b	0.14 NS		

[†] Both significant findings and non-significant positive trends included in this column

* Indicates a significant difference found

NS Indicates no significant difference

Table 1 (continued)

Effects of Tests on Retention Learning: Meta-analysis of Eight Experimental Studies.

	Findings and Sources	Effect size Δ	Mean effect size	Number of positive findings[†]
6	Take-home tests only support retention of material actually appearing on the tests		0.47	2 of 2
	1991	0.17 NS		
	2003b	0.76 *		
7	Short-answer tests support retention learning		0.66	1 of 1
	1994	0.66 *		
8	Post-test reviews support retention learning			1 of 1
	1995	0.34 *		
9	Matching tests support retention learning		0.90	1 of 1
	2003a	0.90 *		
10	Study questions support retention learning		0.85	1 of 1
	2003b	0.85 *		
†	Both significant findings and non-significant positive trends included in this column			
*	Indicates a significant difference found			
NS	Indicates no significant difference			

1997a). This seems counter intuitive because one would assume that students study more diligently when they expect a test than when they do not—the anticipation of an upcoming test would logically drive students to study. It may well be true that immediately after instruction occurred, the groups who expected a test would have had more immediate knowledge in short term memory (having “studied up” for the expected tests). But that was not the point of these experiments; this research concerned delayed retention (defined as learning lasting 3 weeks after instruction). Though there was a very small amount of positive trend favoring the groups who thought they would be tested over those who knew they would not be tested, it appears that no significant amount of meaningful learning was achieved merely because of the anticipation of a test. In both of these studies, however, Factor 2 clearly shows that students who actually did take the announced test did retain significantly more information following the three-week delay period. So, the mere threat of an upcoming test does not increase retention learning unless a test is actually administered.

Factor 2: Does taking a test increase retention learning?

This was the most important factor of consideration in all of the experiments examined. Six of the eight experiments had significant positive results of varying magnitude supporting the premise that taking a test increases retention learning. In the other two studies there were non-significant trends supporting this premise as well. It appears that, regardless of the type of tests used, the act of taking a test helps move information from short term memory to a deeper level. Whether this effect is caused by the mere fact that taking a test provides one additional opportunity for rehearsal or there is some unknown factor (such as the kinesthetic act of writing the answers) at work was not determined by this series of studies and that may be a fruitful topic for further investigations.

Factor 3: Does time on task (including tests) increase retention learning?

This showed positive significant findings in all but one of the studies (Haynie, 2004) and logically follows the findings on Factor 2. In addition to test taking (any type of test), both reviews and use of study questions were shown to aid in retention. The one study that failed to have significant findings on this issue did show a supportive non-significant trend. This finding is in harmony with those in a broad spectrum of the research literature in education—time on task does support both short term and retention learning.

Factor 4: Do take-home tests support retention learning?

The findings on this factor were mixed. The 1991 study had a high positive and significant finding while the 2003b study only supported take-home tests with a very slight non-significant trend (nearly neutral). These research questions were answered on the basis of the total 30-item delayed retention test results. The two subscales within the tests (previously tested information, and novel information) were used to investigate Factors 5 and 6.

Factor 5: Do take-home tests support retention better than in-class tests?

The 1997a study included a simple survey in which students claimed that they prefer take-home tests (80%) but they admitted that other types of tests were more accurate for evaluation (77%). Some students say they learn more with take-home tests but that claim was not fully supported in these experiments. Both of the studies that examined this question (1991 and 2003b) had slightly positive non-significant trends in support of take-home tests over in-class tests on the whole. But closer examination of the findings in these studies on the subtests of previously tested and novel information showed that these gains were entirely due to higher performance on the previously tested information while their performance was actually lower on the novel information. To fully sort out the meaning of this finding Factor 6 must also be examined.

Factor 6: Do take-home tests only support retention of material actually appearing on the tests?

The findings in both of these studies on the subtest of novel information (information that was not reflected on either the in-class or take-home versions of the immediate test) showed that the in-class groups outperformed the take-home groups. In the 1991 study, this was only a non-significant positive trend, but the 2003b study had a clear positive significant finding showing that the groups who were tested in-class had studied the material more fully while the take-home groups had apparently merely hunted for the answers to the specific questions appearing on their take-home tests.

Factor 7: Do short-answer tests support retention learning?

The only study examining this factor (1994) had a positive significant finding. In that study, however, the groups who took multiple-choice tests scored even (significantly) higher than the short-answer test groups.

Factor 8: Do post-test reviews support retention learning?

Only the 1995 study asked this question and it had a significant positive finding. Teachers who invest the time to return tests and review them with students provide additional time on task in addition to the reinforcement value this practice affords. The following quotation from the 1995 study set the stage for this investigation:

One aspect of testing which has received little attention in the literature is the usefulness of post-test reviews. After a test is administered and scored, the instructor typically returns the test to students or provides knowledge of results so that students can see their progress in the course. If the test is actually returned, students may be allowed to ask questions about items they missed, the instructor may review the entire test, or no additional information may be given. Reviewing the entire test is time-consuming, but it may be a justifiable use of class time if significant retention learning occurs due to the review. Alternatively, if retention learning is not enhanced by post-test reviews, then they are a waste of valuable class time. Some researchers have questioned the value of in-class tests and asserted that assignments done at home and other sorts of time on-task work could be equally or more beneficial in promoting learning (Faw & Waller, 1976; Haynie, 1990[b]; Gay & Gallagher, 1976). For those who hold this view, post-test reviews would seem an inexcusable waste of valuable learning time. (Haynie, 1995, p. 80)

The gain shown in this study was beyond the gains already documented from anticipating and then taking a test because the Effect size of .34 was found by contrasting the group that took a test and then had the review against the group that only took the test. By reexamining Factor 2 (in Table 1) it may be seen that both of these groups had already outscored the no-test control group by a high significant Effect size of 1.29. It would not be legitimate to combine these two Effect sizes by merely adding them, but it is clear that if taking the test results in an Effect size of 1.29 and then there is a further increase due to the

post-test reviews as documented by the .34 Effect size favoring the review group, the combined benefits of a test followed by a review are considerable.

Factor 9: Do matching tests support retention learning?

Yes, the 2003a experiment had a significant positive finding which showed that matching tests do support retention of the information tested. In fact, in this single study, there was even a small significant difference in the scores of the two tested groups which favored the matching test group over the multiple-choice test group, but the actual difference in the means of the scores was so small that it would be of no practical significance. No claim for superiority of matching tests was made in the conclusions of this study—both matching and multiple-choice tests were reported to support retention learning.

Factor 10: Do study questions support retention learning?

The study that considered this question used two groups with the exact same treatment but with different names. In the 2003b study, one of the groups took the take-home test and another group was given the same exact set of questions with the heading “Study Questions.” They were not told that other groups had the same information as was on the take-home test. When tested three weeks later, the group with the study questions outperformed all of the other groups (even the in-class test group) despite the fact that they were not actually tested. It is presumed that these students, unlike the take-home test students, did in fact read the entire study booklet, studied it broadly, and then used the study questions as an aid to further review and study. They did expect that a test was forthcoming, but they were not actually tested, so they likely prepared as well as the in-class test group or (apparently) better.

Conclusions and Recommendations

Ten factors were examined in this series of eight related experiments. The methods of the studies were similar except for the treatments related to immediate testing. The dependent variable in all of the studies was a delayed retention test that was common to all groups in all of the studies, enabling clear comparisons among the studies. The instructional materials and tests all concerned technical subject matter about “high tech” materials used by NASA and their applications beyond the space program. The tests went beyond mere memorization of facts to also represent levels 2 and 3 of Bloom’s taxonomy: Comprehension and Application. All ten of the factors were found to have some supportive findings, though for two of them the support was merely non-significant trends in each. The remaining 8 factors all had at least one study with a significant supportive finding. No negative findings were present. The ten factors and the findings related to them are discussed fully in the preceding section of this article.

The most persuasive evidence among the eight studies was support for the hypotheses that taking a test on material studied and increased time on task (whether in the form of a test or other activities such as reviews or use of study questions) both result in increased retention learning by students. Even given the

hands-on nature of much of the learning in technology education courses, the cognitive learning is best assessed with traditional classroom tests and it is important to maximize the learning value of the time spent in testing. Further, delayed retention learning, as evaluated in these experiments, is of far more value than the short term recall evidenced when students take a test for which they have recently “crammed.” If all of the findings of this meta-analysis and the individual studies are considered together, it would appear that the best practice for increasing retention learning of students would be a well orchestrated protocol in which:

1. An upcoming in-class test is announced at the time the unit of study commences.
2. Study questions reflecting about 2/3 of the test are provided. The study questions should be alternate versions of the actual test items they reflect.
3. The test should be administered as promised.
4. In the first class session following test administration, the test should be returned after it has been graded and a thorough post-test review should be conducted in which students see their scores and are allowed to ask follow-up questions about items that they missed.

This recommended procedure does require a lot of class time and diligence by the teacher to grade tests as soon as possible after they are given. It is also understood that some students may become alienated or even argumentative if they feel that ambiguous or “tricky” items have harmed their score and their eventual course grade. Though such moments would be uncomfortable, the prudent teacher will then follow the recommendations in previous works (Haynie, 1983, 1992, and 1997b, or any text on test construction) to improve the weak or flawed items for the benefit of future students. These and other previous studies have shown that teacher-made classroom tests, though valuable for many reasons, often have serious flaws. Only when well prepared tests are administered properly, graded quickly, and reviewed effectively will the maximum gains in retention learning be achieved by students. This is a large investment of time and effort by both teachers and students, but if learning is not aided by testing, the testing itself is a waste of resources and time. Only important cognitive learning should be treated in this thorough manner, but if facts, concepts, or abilities to apply learning to novel situations are truly important for accomplishing technology course objectives, this holistic approach will enhance the likelihood of students retaining what they learn.

Future studies in this vein should examine questions related to instruction and testing via computers, testing issues in distance learning settings, and follow-up investigations to determine what it is about taking a test that supports retention. Perhaps there are ways to further enhance those particular elements or actions that support the retention learning gains more efficiently. Testing will remain a value-charged issue worthy of future research. At present there is a trend toward the evaluation of many more learning products and activities via

rubrics and other means that draw attention away from traditional tests. Nonetheless, classroom tests will remain a very prominent feature of education (including technology education) for the foreseeable future and educators should invest the time to use them well.

References

- Borg, W. R., & Gall, M. D. (1989). *Educational Research: An Introduction*. 5th ed. New York: Longman.
- Burdin, J. L. (1982). Teacher certification. In H. E. Mitzel (Ed.), *Encyclopedia of education research* (5th ed.) New York: Free Press.
- Carter, K. (1984). Do teachers understand the principles for writing tests? *Journal of Teacher Education*, 35(6), 57-60.
- Denny, J. D., Paterson, G. R., & Feldhusen, J. F. (1964). Anxiety and achievement as functions of daily testing. *Journal of Educational Measurement*, 1, 143-147.
- Duchastel, P. (1981). Retention of prose following testing with different types of test. *Contemporary Educational Psychology*, 6, 217-226.
- Faw, H. W., & Waller, T. G. (1976). Mathemagenic behaviors and efficiency in learning from prose. *Review of Educational Research*, 46, 691-720.
- Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway (Ed.), *Testing in the schools: New directions for testing and measurement*, No. 19 (pp.29-38). San Francisco: Jossey-Bass.
- Gay, L., & Gallagher, P., (1976). The comparative effectiveness of tests versus written exercises. *Journal of Educational Research*, 70, 59-61.
- Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. *Educational Measurement: Issues and Practice*, 4(1), 15-18.
- Haynie, W. J. (2004). Effects of pre-tests and post tests on delayed retention learning in technology education. *North Carolina Journal of Technology Teacher Education*, VI, 14-21.
- Haynie, W. J. (2003a). Effects of multiple-choice and matching tests on delayed retention learning. *Journal of Industrial Teacher Education*, 40(2), 7-22.
- Haynie, W. J. (2003b). Effects of take-home tests and study questions on delayed retention learning. *Journal of Technology Education*, 14(2), 6-18.
- Haynie, W. J. (1997a). Effects of anticipation of tests on delayed retention learning. *Journal of Technology Education*, 9(1), 20-46.
- Haynie, W. J. (1997b). An analysis of tests authored by technology education teachers. *Journal of the North Carolina Council of Technology Teacher Education*, 2(1), 1-15.
- Haynie, W. J. (1995). In-class tests and posttest reviews: Effects on delayed-retention learning. *North Carolina Journal of Teacher Education*. 8 (1), 78-93.

- Haynie, W. J. (1994). Effects of multiple-choice and short-answer tests on delayed retention learning acquired via individualized, self-paced instructional texts. *Journal of Technology Education*, 6(1), 32-44.
- Haynie, W. J. (1992). Post hoc analysis of test items written by technology education teachers. *Journal of Technology Education*, 4(1), 27-40.
- Haynie, W. J. (1991). Effects of take-home and in-class tests on delayed retention learning acquired via individualized, self-paced instructional texts. *Journal of Industrial Teacher Education*, 28(4), 52-63.
- Haynie, W. J. (1990a). Anticipation of tests and open space laboratories as learning variables in technology education. *Journal of the North Carolina Council of Technology Teacher Education*, 1(1), 2-19.
- Haynie, W. J. (1990b). Effects of tests and anticipation of tests on learning via videotaped materials. *Journal of Industrial Teacher Education*, 27(4), 18-30.
- Haynie, W. J. (1983). Student evaluation: The teacher's most difficult job. *Monograph Series of the Virginia Industrial Arts Teacher Education Council*, Monograph Number 11.
- Herman, J., & Dorr-Bremme, D. W. (1982). *Assessing students: Teachers' routine practices and reasoning*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Marsh, R. (1984). A comparison of take-home versus in-class exams. *Journal of Educational Research*, 78(2), 111-113.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87-99.
- Mehrens, W. A. (1987). "Educational Tests: Blessing or Curse?" Unpublished manuscript, 1987.
- Mehrens, W. A., & Lehmann, I. J. (1987). Using teacher-made measurement devices. *NASSP Bulletin*, 71(496), 36-44.
- Newman, D. C., & Stallings, W. M. (1982). *Teacher Competency in Classroom Testing, Measurement Preparation, and Classroom Testing Practices*. Paper presented at the Annual Meeting of the National Council on measurement in Education, March. (In Mehrens & Lehmann, 1987).
- Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, 74(1), 18-22.
- Stiggins, R. J., & Bridgeford, N. J., (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286.
- Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, 5(2), 5-17.
- Thorndike, R. L., & Hagen, E. P. (1977). *Measurement and evaluation in psychology and education*. New York: Wiley.