U. S. NUCLEAR REGULATORY COMMISSION
REGION I

REPORT NUMBER:	50-245/92-32 (OL)
DOCKET NUMBER:	50-245
LICENSE NUMBER:	DPR-21
LICENSEE:	Northeast Nuclear Energy Company P.O Box 270 Hartford, CT 06141-0270
FACILITY:	Millstone Unit 1 Nuclear Power Station
INSPECTION DATES:	December 14 - 17, 1992
ACCOMPANIED BY:	J. H. Williams, Sr. Operations Engineer J. S. Stewart. Operations Engineer M. Parrish, EG&G

LEAD INSPECTOR/ EXAMINER: Donald J. Florek 12/29/92
Donald Florek, Sr. Operations Engineer Date
BWR Section, Operations Branch
Division of Reactor Safety

APPROVED BY: Todd [signature] for 12/29/92
Richard J. Conte, Chief Date
BWR Section, Operations Branch
Division of Reactor Safety

INSPECTION SUMMARY: Examination/Inspection conducted December 14 - 17, 1992 (Inspection Report 50-245/92-32)(OL)

AREAS INSPECTED: Routine announced examination/inspection to administer retake requalification examinations to two operators and three senior operators. The inspection reviewed facility licensee actions on previous inspection findings related to emergency operating procedures, independent audits of operator training programs and status of selective corrective actions from December 3, 1992, meeting to correct performance problems identified during the requalification examination period.

RESULTS: Two operators and two senior operators passed the examination. One senior operator failed in use of the emergency plan as part of the simulator evaluation. Unresolved item (URI 245/89-19-01) regarding emergency operating procedures was closed. A new unresolved item (URI 245/92-32-01) was opened for issuance of flowchart based emergency operating procedures. The performance of the operator training audits required by ANSI N18.7 and Technical Specification 6.5.3.7(b) is unresolved (URI 245/92-32-02). The facility licensee is proceeding on with the specific commitments made at the December 3, 1992, meeting. The effectiveness of the licensee actions can only be assessed based on future observations after more data is obtained.

DETAILS

1.0 INTRODUCTION

Examination report 50-245/92-23 (OL) documented the unsatisfactory Millstone Unit 1 requalification program evaluation. Following the unsatisfactory program evaluation, a Confirmatory Action Letter (CAL) 1-92-014 was issued that restricted those licensed operators who failed either the NRC or facility requalification examination until either the facility requalification program was determined to be at least provisionally satisfactory or until the licensed operators have been evaluated as successful by the NRC. During the week of December 14, 1992, the NRC staff administered requalification retake examinations to 4 licensed operators (2 ROs and 2 SROs) who had failed NRC administered requalification examinations and one SRO who failed a facility requalification examination. One RO and two SROs were administered a simulator examination, one RO was administered a JPM examination and one SRO was administered JPMs to evaluate his ability to classify an event. The JPM methodology was used in lieu of classification in a dynamic simulator environment. The examiners used the process and criteria described in NUREG-1021, "Operator Licensing Examiner Standards," Revision 6. A listing of the items used in the examination is contained in Attachment 1.

2.0 SUMMARY OF EXAMINATION

2.1 Individual Examination Results

NRC GRADING

	RO Pass/Fail	SRO Pass/Fail	TOTAL Pass/Fail
Written	N/A	N/A	N/A
Simulator	1/0	2/1	3/1
Walk-through	1/0	N/A	1/0
Overall	2/0	2/1	4/1

2.2 Strengths and Weaknesses Based on Individual Operator Performance

No generic weakness was noted. The examiners noted that the teamwork of the crew examined in the scenario has significantly improved from prior examinations.

2.3 Simulator Remedial Training Program

The facility licensee had initially planned on the NRC reexamining two reactor operators on the simulator who had failed the prior NRC administered requalification. Millstone Unit 1 operates with three reactor operators in the RO, BOP and RPO position so an operator who had previously passed an NRC examination would also be used. The two scenarios developed for the examination resulted in critical tasks for the RO and BOP position but no critical tasks for the RPO position. The NRC planned to rotate the two reactor operators into the RO and BOP position and use the third operator in the RPO position that had no critical tasks to perform.

After the facility licensee had observed the validation of the two simulator scenarios, they requested that one of the reactor operators not rotate into the BOP position. The basis for the facility request was that the RO had not practiced in the BOP position during the remediation training program since no weakness was identified for the individual while in the BOP position. The facility licensee indicated that while they had confidence that the individual could perform in the BOP position they did not want to subject the individual and crew to an examination situation that they were not prepared for. The other reactor operator being examined operated in all panel positions during the remediation program. The facility licensee requested an additional day of training and evaluation for the crew in the proposed configuration. The NRC examiner indicated that additional training was inappropriate. On the morning of the examination, the facility licensee elected to substitute another licensed operator from the training staff who worked with the crew during the remediation period to complete the crew. The facility licensee requested that two short 5-10 minute warmup scenarios with the additional substitute operator. The NRC examiner was present at the warmup scenarios and determined that no compromise of the examination occurred.

2.4 Emergency Plan Classification JPM Results

The SRO who was administered JPMs to evaluate his ability to classify an event did not pass the retake examination. The JPM methodology was used in lieu of classification in a dynamic simulator environment. Two JPM scenarios were used. One was selected from the bank and the other was modified from the bank.

Following the examination results, the examiner reviewed the activities of the remediation period. The SRO had asked for and was provided a copy of the JPM scenarios in the bank (12 scenarios) for emergency classification. Since the facility licensee uses the bank for quarterly take-home-then-send-back-results testing of SROs, which was ongoing, the facility licensee did not provide the answers to the SRO. The examiner reviewed a copy of the

SRO's personal bank with the answers that were filled in. Two of the classifications were in error; one of these was selected for the examination. The SRO determined the emergency classification in the examination to be the same as in his personal copy. At the end of the examination week, the facility licensee had not completed their review to determine if the SRO or training staff filled in the answers for the SRO's personal bank.

The examiner was concerned about the release of the emergency plan classification bank when only 12 scenarios were available in the bank. The size of the emergency plan classification bank is not sufficiently large to release to preclude examination preconditioning of the operators. At the exit meeting, the facility licensee representatives acknowledged that additional controls were needed for the JPMs that contain scenarios for emergency classification and agreed to review this area.

3.0 LICENSEE ACTIONS ON PREVIOUS INSPECTION FINDINGS

(Closed) Unresolved Item 245/89-19-01 Adequacy of three technical deviations from the BWR Owners Group Emergency Procedure Guidelines regarding use of the isolation condenser, initiation of primary containment flooding and emergency depressurization to protect primary containment and secondary containment.

Inspection Report 50-245/92-11 (OL) indicated that the additional facility licensee information provided in a letter dated February 10, 1992, required further evaluation. NRC Letter dated July 17, 1992, resolved the outstanding technical issue regarding the three technical deviations and indicated that the related procedure changes will be assessed in future inspections of licensed activities. The NRC inspector reviewed current draft versions of the EOP flowcharts and determined that the procedure changes required were contained in the draft flowcharts. The licensee has indicated in a letter dated December 2, 1992, that the implementation of flowchart based EOP will not be completed until the end of July 1993. Pending issuance of the flowchart EOPs and review by NRC staff, this area is unresolved. (URI 245/92-32-01)

4.0 INDEPENDENT AUDITS OF OPERATOR TRAINING PROGRAMS

The inspector discussed quality assurance and other licensee-performed independent audits of operator training programs with the licensee. The inspector also reviewed how the technical specifications audit requirements of paragraph 6.5.3.7 were being met. The licensee's QA plan commits to ANSI N18.7-1976 which requires audits of programs for training, retraining, qualification and performance of operating staff. Technical Specification 6.5.3.7(b) requires audits be performed of training of the unit staff at least once per 12 months. In the period of 1986 to 1992, there were only two narrowly focused audits associated with operator training identified. Both audits were performed in 1992. One audit of class attendance was requested by the Training Department and the other dealt with lesson plan modification.

This inspection is incomplete because of the operator requalification retake examination efforts. Therefore, more information and discussion with the licensee is needed to resolve the issue of auditing training programs of the unit staff. The performance of the operator training audits required by ANSI N18.7 and Technical Specification 6.5.3.7(b) is unresolved (URI 245/92-32-02).

5.0 MANAGEMENT MEETING ON DECEMBER 3, 1992, AND INSPECTION FOLLOW-UP

On December 3, 1992, a management meeting was held at the Region 1 offices to allow the facility licensee to present the facility analysis of why the 1991 remediation effort failed to prevent the 1992 weaknesses and the corrective actions that the facility licensee will take. Attachment 3 contains a listing of the individuals in attendance at the meeting. The facility licensee handouts provided at the meeting are contained in Attachment 4. The facility formally submitted the short and long term corrective actions and the status of previously committed long term corrective actions in a letter dater December 7, 1992.

During the site visit from December 13 - 17, 1992, the inspector also reviewed and discussed with facility licensee personnel some of the corrective actions identified at the December 3, 1992, meeting. The inspector reviewed the facility licensee activities in criterion referenced testing, advanced training classroom, quantifying simulator performance, use of Vermont Yankee simulator, use of split page format for instructional materials, spacing effect, interleaving, instructor crew accountability and visualization. The effectiveness of most of the items cannot be assessed until they have been used more and additional information is obtained.

Criterion referenced testing was used to identify generic weakness by testing 25% of the licensed operators using an extensive series of tests. When 30% or more of the operators examined answered incorrectly, a generic weakness was identified. Electrical, instrumentation logic, recirculation system, isolation condenser and control rod drive were the major generic system weaknesses identified. Purportedly, criterion referenced testing effectively eliminates guessing and encourages operators, in a nonthreatening manner, to admit or demonstrate a knowledge weakness. The facility licensee is also planning to apply the methodology to pre-test areas for future lessons to adjust the training for the specific knowledge weaknesses of operators to be trained. The examiner concluded that criterion referenced testing appeared to be a good tool to provide feedback to the trainers for the areas that operators need remediation training.

The facility licensee has set up two classrooms using a computer based individual student feedback system on the material to be covered. The lesson plans have preestablished questions in the lesson to ask the students, as they are being trained, to get prompt feedback on whether the concept is understood by all students. Individual, as well as group, data is obtained. The instructor can easily identify if the students as a group or individually are having difficulty understanding the lesson being taught. Because entire class feedback is

required, student attentiveness is also increased. The examiner concluded that this type of classroom setup also appears to have a high potential to improve student learning.

The facility licensee also has setup a computer based system to try to measure student improvement in the simulator. When the students are being evaluated, instructors quantitatively evaluate communications, analytical skills, technical performance and teamwork at predetermined intervals. Superimposed on the quantitative evaluations are verbal comments of the evaluators as feedback to the individuals being evaluated. This system has recently been installed.

Facility licensee evaluation of the use of the Vermont Yankee simulator as part of teamwork enhancement training has been positive. The training reinforced working together to understand plant response because of the unfamiliarity of the VY control panel and procedures.

The facility licensee is investing resources in using split page training material, spacing and interleaving of training presentations and visualization methodology to enhance student learning and retention. These methods are founded in various research studies to improve student learning and retention. The effects of these methods are pending after more data is obtained on operator performance.

The facility licensee is establishing an instructor/operator accountability by assigning a specific instructor for each crew. The instructor will be responsible for assuring that each member of the crew is performing up to expectations.

The facility licensee is proceeding with the specific commitments made at the December 3, 1992, meeting. The effectiveness of the licensee actions can only be assessed based on future observations after more data is obtained.

6.0 EXIT MEETING

Region I conducted an exit meeting on December 17, 1992. The NRC presented the results of the examinations and discussed inspection related findings as discussed in this report. Those persons in attendance are noted in Attachment 2.

ATTACHMENT 1

REQUALIFICATION EXAMINATION TEST ITEMS

JOB PERFORMANCE MEASURE'S USED

61 Classify an Event Using Emergency Plan Implementing Procedures
14 Manual Start from Diesel Room
18 Transfer Reactor Protection System Bus A to Reactor Protection Motor Generator A
83 Initiate Drywell Cooling
12 Isolate Control Rod Drive 30-39
113 Standby Liquid Control Pump Operability Test (Faulted)
24 Transfer to Mechanical Pressure Regulator at Power
31 Core Spray Operability Test
25 Recirc Pump Startup
1 Return Isolation Conder to Standby
54 Restore power to res ve startup transformer after loss of normal power

SIMULATOR SCENARIOS

1. Failure of the reactor protection system to automatically initiate a scram with an increasing scram discharge volume water level. The B battery fails following the manual scram resulting in loss of about 50% of the annunciators. An un-isolatable leak in reactor water cleanup then occurs.

2. Control rod drive pump A is out of service. The diesel generator fails during a surveillance. A loss of power occurs and the control rods do not insert into the core. The B standby liquid control (SLC) pump trips resulting in inability to either manually drive rods or initiate SLC. When operators cross connect the gas turbine to the diesel bus, operators will be able to manually drive control rods and inject baron with SLC.

ATTACHMENT 2

PERSONS CONTACTED

NORTHEAST UTILITIES

*H. Haynes, Director Unit 1
*R. Heidecker, Supervisor Operator Training Unit 1
F. Libby, Quality Services Department
*P. Przekop, Operations Manager
*M. Ross, Operations Assistant
*B. Ruth, Manager Operator Training
A. Saunders, Quality Services Department

NUCLEAR REGULATORY COMMISSION

*D. Florek, Sr. Operations Engineer
*K. Kolaczyk, Resident Inspector
*R. Laufer, Examiner NRR/OLB
*H. Williams, Sr. Operations Engineer

*Denotes those present at the exit meeting on December 17, 1992

ATTACHMENT 3

PERSONS IN ATTENDANCE AT DECEMBER 3, 1992, MEETING

NORTHEAST UTILITIES

M. Black, Director Nuclear Training
M. Brown, Project Manager Operator Training Unit 1
H. Hayr Director Unit 1
R. Heidecker, Supervisor Operator Training Unit 1
P. Przekop, Operations Manager
W. Romberg, Vice President Nuclear Operations Services
M. Ross, Operations Assistant
M. Wilson, Supervisor Nuclear Licensing

NUCLEAR REGULATORY COMMISSION

J. Anderson, Project Manager, NRR
L. Bettenhausen, Chief Operations Branch
P. Bissett, Sr. Operations Engineer
L. Briggs, Sr. Operations Engineer
R. Conte, Chief BWR Section
L. Doerflein, Chief Reactor Project Section 4A
D. Florek, Sr. Operations Engineer
W. Hodges, Director Division of Reactor Safety
T. Martin, Regional Administrator
J. Munro, Sr. Operations Engineer, NRR
R. Pelton, Training Specialist, NRR
C. Sisco, Operations Engineer
J. Stewart, Operations Engineer
H. Williams, Sr. Operations Engineer

ATTACHMENT 4

FACILITY HANDOUTS AT THE DECEMBER 3, 1992, MEETING

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Northeast Nuclear Energy Company
Millstone Unit No. 1

Licensed Operator Requalification Program Presentation

December 3rd 1992

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Wayne Romberg

Vice President - Nuclear

Operations Services

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Briefing Outline

Malcolm Black

Director Nuclear Training

milistone
Unit 1 Nuclear Power Station

NRC BRIEFING
KING OF PRUSSIA
DECEMBER 3RD, 1992

EXCELLENCE IN OPERATIONS

PROGRAM OUTLINE

° INTRODUCTION	WAYNE ROMBERG
° BRIEFING OUTLINE	MALCOLM BLACK
° PROJECT TEAM INTRODUCTION	MALCOLM BLACK HARRY HAYNES
° CURRENT STATUS OF MILLSTONE UNITS 2, 3, AND HADDAM NECK PLANT TRAINING PROGRAMS	MALCOLM BLACK
° EXPLANATION OF WHY THE 1991 REMEDIATION EFFORT FAILED TO PREVENT THE 1992 WEAKNESSES	BRAD RUTH
° TRAINING PROGRAM FAILURE ANALYSES	ROBERT HEIDECKER
° TRAINING STRATEGIES AND PROGRESS REPORT	MICHAEL BROWN
° PARTNERSHIP EFFORT	PETE PRZEKOP
° STRATEGIC RECOVERY PLAN - SHORT AND LONG TERM	ROBERT HEIDECKER
° CLOSING	WAYNE ROMBERG

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

The

Project Team Introduction

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Malcolm Black - Director
Nuclear Training

Harry Haynes - Director
Millstone Unit 1

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Mike Brown

Project Position: Project Manager

Recent Assignment: Manager, Technical Training - 6 yrs (managed the initial accreditation of NUs' Technical Training Programs)

Experience:
- I&C Manager, Millstone Unit 3 - 9 years (Design thru commerical operations)
- Head up I&C startup of Millstone 2 CPF - 2 years
- Millstone 2 I&C - 3 years (start up)
- Millstone 2 Operations Dept. - 1 year
- Navy Nuclear Program - 6 years (Reactor Operator)

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Robert Heidecker

Project Position: Supervisor, Operator Training

Recent Assignment: Supervisor, Operator Training, Haddam Neck past 5 years

Experience:
- o 4 yrs consultant at Haddam Neck Training
- o 6 yrs training/supervisory experience at Westinghouse
- o 12 yrs Navy Nuclear experience

Educational Experience/Licenses: BS Degree - Nuclear Engineer, Penn State
SRO License - Beaver Valley Unit 1
SRO Certified (NRC) - Haddam Neck

Professional Associations: Representative, BWR/PWR owners group - simulator complexity/critical tasks

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Burns Mixon

Project Position: Human Performance Specialist

Recent Assignment: Consultant to Washington Public Power and Commonwealth Edison's (NU employee as of 12/15/92) recovery effort

Experience:

- Commonwealth Edison & Washington Public Power Human Performance consultant focusing on leadership, teamwork, decision making, and communciation skills - 1989 - present.
- Human Resource consultant to industry other than nuclear-related - 14 years.
- Faculty U.S. Air Force Academy - 7 years (lecturing on leadership topics).
- BA and MA - Florida State University - 1965 and 1971

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Peter Przekop

Project Position: Operations Manager

Recent Assignment: I&C Manager, Millstone Unit 1 - 7 years

Experience:

- o SRO licensed MP-1 - 18 years
- o Unit 1 Duty Officer
- o New Haven College 1970 BS Physics
- o Florida State University Physics
- o Staff Assistant-Millstone Unit 1 Superintendent - 3 years
- o Staff Assistant-Millstone Station Superintendent - 2 years
- o Engineering Supervisor, Millstone Unit 1 - 2 years
- o Engineer, Millstone Unit 1 - 3 years
- o Maintenance Engineer, Millstone Unit 1 - 3 years

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Mike Ross

Project Position: Operations Department Liaison

Recent Assignment: Operations Assistant past 4 years

Experience:
- SRO Licensed
- Millstone Unit 1 Duty Officer
- University of Connecticut BME
- Operations Shift Engineer-2 years
- Unit 1 Engineering (Qualified Reactor Engineer) 5 years

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

CURRENT STATUS

OF

MILLSTONE UNITS 2,3 AND HADDAM NECK

REQUALIFICATION TRAINING PROGRAMS

millstone
Unit 1 Nuclear Power Station

MILLSTONE UNITS 2,3 AND HADDAM NECK

REQUALIFICATION PROGRAM ASSESSMENTS

- o NRC REQUALIFICATION EXAMINATIONS
  - -49 OF 50 PASSED (1991 TO PRESENT)

- o NRC REQUALIFICATION TRAINING INSPECTION
  - -MEET 10CRF55 REQUIREMENTS

- o NU ASSESSMENTS USING NUREG - 1220 CRITERIA
  - -QUALITY ASSESSED AS SATISFACTORY

- o NU BIENNIAL ACCREDIATION SELF ASSESSMENT
  - -PROGRAMS MEET ACCREDITATION OBJECTIVES

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

CURRENT STATUS

- MILLSTONE UNITS 2,3 AND HADDAM NECK REQUALIFICATION PROGRAMS ARE SATISFACTORY

- SOME ACTIONS ARE NECESSARY

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

NU ACTIONS

IN PROGRESS

- o PERFORMANCE ENHANCEMENT PROGRAM
  - INCREASE AND RETAIN QUALIFIED STAFF
  - STAFF AUGMENTED WITH CONTRACT INSTRUCTORS
  - TRAINING MATERIAL IMPROVEMENTS
  - STAFFING CONTINGENCY FOR UNPLANNED ATTRITION SITUATIONS
- o TRAINING MANAGEMENT SYSTEM UPGRADE
- o OPERATOR TRAINING BRANCH ADMINISTRATIVE PROCEDURE IMPROVEMENTS

PLANNED

- o EVALUATE AND IMPLEMENT A MORE RIGOROUS SELF ASSESSMENT PROCESS

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

WHY THE 1991 REQUALIFICATION PROGRAM AND CORRECTIVE ACTIONS FAILED TO PREVENT 1992 WEAKNESSES

millstone

Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER

NU

EXCELLENCE IN OPERATIONS

1992 REQUALIFICATION OPERATOR PERFORMANCE SUMMARY

1991 ASSESSMENTS	1992 ACTIONS	RESULTS
<u>STRENGTHS</u>	ADVANCED SCENARIOS INCORPORATED	
WRITTEN EXAMINATION PERFORMANCE	OPERATIONS AND STAFF CREWS TRAINED TO NEW COMMAND AND CONTROL STANDARDS	WRITTEN EXAMINATION PERFORMANCE DECLINED
JPM PERFORMANCE		JPM PERFORMANCE DECLINED
<u>WEAKNESSES</u>	EOP IMPLEMENTATION AND MONITORING OF KEY PARAMETERS STRESSED	
SRO COMMAND, CONTROL, AND COMMUNICATIONS	LICENSE HOLDERS FAILING 1991 EXAM REMEDIATED AND RE-EXAMINED	CREW COMMAND, CONTROL AND COMMUNICATIONS IMPROVED
CREW TEAMWORK		
EOP IMPLEMENTATION	ADDITIONAL TEAMWORK TRAINING	SPECIFIC EOP WEAKNESSES CORRECTED
CREW MONITORING OF KEY PARAMETERS	FREQUENCY OF CREW EVALUATIONS INCREASED	OVERALL <u>CREW</u> ATTENTION TO CRITICAL PARAMETERS AND SAFETY SYSTEM ACTUATIONS IMPROVED
	SIMULATOR CHANGES IMPLEMENTED	

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

1992 REQUALIFICATION TRAINING PROCESS SUMMARY

1991 ASSESSMENTS	1992 ACTIONS	RESULTS
<u>WEAKNESSES</u> EVALUATOR OBJECTIVITY EXAMINATION MATERIALS	SIMULATOR EVALUATION TOOLS REVISED TO REFLECT NEW STANDARDS EVALUATION TEAM OBSERVED AND FEEDBACK PROVIDED ADVANCED SCENARIOS INCLUDED IN EXAMINATION BANK JPM MEASUREMENT STANDARDS REVISED JPM QUESTIONS NOT ADDRESSED OPEN REFERENCE QUESTIONS REVISED	EVALUATORS OBJECTIVE JPM MEASUREMENT STANDARDS ACCEPTABLE JPM QUESTIONS REQUIRED REVISION FOR EXAMINATION REVISED OPEN REFERENCE QUESTIONS NOT WRITTEN TO HIGHER COGNITIVE LEVELS JPM PROCESS PROBLEMS WERE IDENTIFIED

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

CAUSES FOR 1992 WEAKNESSES

- NU SELF ASSESSMENT FOCUSED ON OBSERVED WEAKNESSES

- LOWER COGNITIVE LEVEL OF EXAMINATION BANK ITEMS FAILED TO EXPOSE EXISTING OPERATOR KNOWLEDGE DEFICIENCIES

- 1992 REQUALIFICATION TRAINING CENTERED ON THE SIMULATOR EMPHASIZING "CREW SKILLS" AND "EOP IMPLEMENTATION SKILLS"

- JPM TRAINING WAS NOT FORMALLY SCHEDULED INTO THE REQUALIFICATION TRAINING PROGRAM

- EXTENDED PERIODS WITH REDUCED NUMBERS OF TRAINING STAFF AND LOW EXPERIENCE LEVELS CONTRIBUTED TO PROGRAM DEGRADATION

millstone
Unit 1 Nuclear Power Station

ANNUAL EXAMINATION RESULTS

Simulator Examinations

- o 86% pass rate
- o One crew failure
- o Three operators failed due to E-plan classification (2 CO's & 1 SCO)
- o Overall crew performance improved

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

EXAMINATION RESULTS

JPM Examinations

- o 97.6% pass rate
- o 19 of 430 JPMs were failed
- o 4 JPMs were missed twice
- o One individual failed (1 RO)

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

EXAMINATION RESULTS

Written Examinations

- o 84% pass rate
- o 7 individuals failed (4 ROs, 1 SS, 2 Staff SRO)
- o Marginal performance compared to other units/industry
- o 65% of scores were greater than 85% (average score = 86.8)

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

SELF ASSESSMENT PROCESS

- o ROOT CAUSE ANALYSIS
- o EXAM ASSESSMENTS
- o CONFIDENCE WEIGHTED TEST

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Self Assessment Analysis

Simulator Weaknesses

- o Simulator training focused primarily at SRO level
- o Limited focus on EOPs versus tasks and system knowledge
- o E-plan classifications not reinforced at crew level
- o Exercise guides did not contain instructional activities/strategies
- o Exercise guides did not contain same standards used during evaluations
- o Simulator time was focused on the EVALUATION mode versus TRAINING mode

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Self Assessment Analysis

JPM Weaknesses

- JPMs not formally scheduled into training program
- Individuals failed to maintain their own proficiency
- Time critical JPMs were not trained on
- Several JPM failures attributed to lack of self checking
- Operators improperly processed initiating cues or initial conditions

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Self Assessment Analysis

Written Weaknesses

- o Exam bank did not adequately assess higher cognitive knowledge levels
- o 40% of exam items tested objectives that were outside the two year sample plan
- o Operators had difficulty processing information due to systems/procedure knowledge deficiencies

millstone
Unit 1 Nuclear Power
Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Programmatic Weaknesses

- o Selected systems knowledge
- o Selected EOP/ONPs
- o E-plan classification for non SS/SCO watchstander

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Self Assessment Analysis

Process Weaknesses

- o JTA and associated training materials being updated
- o Program not designed with proper emphasis in all environments

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Project Goal

To elevate the Millstone Unit 1 Operator Training programs from their current level of marginal acceptability to that of an industry leader

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Strategies to Strengthen:

- o The Training Process
- o The Individual Knowledge Base
- o Performance Skills
- o Teamwork
- o JPM Implementation
- o Program Evaluation

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

The Training Process

- o Reconstructing the Job Analysis for all MP-1 Operator positions
- o Reconfiguring the objectives at the appropriate cognitive levels
- o Upgrading all MP-1 Lesson Plans that have programmatic deficiencies

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

The Training Process

- Upgrading the exam bank to the appropriate cognitive level and congruent with the revised objectives, using the NRC reviewed model

- Reformatting all student text in a scholar text format to enhance learning

- Upgrading student and teaching text graphics to enhance the visual learning element

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

The Individual Knowledge Base

- Capitalize on advanced learning research conducted at NU

-Advanced Training Classroom
-Accelerated Learning
-Examinations as a Preservative of Knowledge
-Harvard (Medical School) study groups "Pathways to the Future"
-Interleafing
-Spacing Effect
-Introduction of flow charts, job aids, and mental cueing to enhance student retention
-Upgrade of the visual component of the student text and classroom presentation materials

millstone
Unit 1 Nuclear Power
Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

The Individual Knowledge Base

- Dr. Robert Bjork will consult monthly on the effectiveness of the learning strategies employed

- Expanded use of interactive video instruction

- Introduction of designer software to teach circuit logics

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Performance Skills

o Enhance Simulator Exercises

- -Greater use of freeze and video critquing techniques
- -Improved facilitation by the Operations Management
- -Implementation of Rev. 7 (a crew success concept)
- -New EOP flow charts

millstone
Unit 1 Nuclear Power Station

Performance Skills

- o Implement the use of mental imagery
- o Develop a model to measure individual and crew performance and improvements over time
- o Improve OJT guides
- o New barrier failure table training aid
- o Expanded use of Technical Training labs and instructors
- o The use of non plant-specific simulators to evaluate crew analytical skills

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Teamwork

- o Additional teamskills training, both individual and crew
- o Increased emphasis on crew evaiuations on the simulator
- o Addition of a Human Performance Specialist to the staff

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

JPM Implementation

- o Upgrade of JPM's
- o Joint unit evaluations of individuals performing JPM's
- o Improved method of implementing JPM's (from a process perspective)

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NNU
EXCELLENCE IN OPERATIONS

Program Evaluation

- o Introduction of confidence weighted testing
- o The use of a curriculum development coordinator to monitor the quality of curriculum improvements
- o Preparation of an INPO self evaluation report
- o Programs will undergo reaccreditation-first quarter 1994
- o NRC initial and requal examinations-1993
- o Evaluation team of various Region 1 utilities
- o Requested INPO Assist Visit - 1993

millstone
Unit 1 Nuclear Power Station

Improvements to Date

- o Upgrade of the NLO text
- o Advance training classroom
- o New model to set proper cognitive level
- o New barrier failure table training aid
- o Technical Trainers and labs into the operator programs
- o Confidence weighted testing
- o Revised JTA and objectives (in progress)

millstone
Unit 1 Nuclear Power Station

Improvements to Date

- o Improved visuals (in progress)
- o Curriculum development coordinator
- o Accelerated learning techniques
- o Programmatic review of JPMs and OJT guides
- o Diagnostic evaluations performed on the individuals who failed this past year
- o Actively pursuing job offers to fili vacant positions

millstone
Unit 1 Nuclear Power Station

Partnership

- o Training initiatives and improvements are only part of the equation
- o Success requires Operations Department input, proactive support, and commitment

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Expectations

- o Training programs must reflect Operation philosophy
- o Standards of operations will be consistent in the Unit and Training environments

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Commitment

o Unit Operations will be an integral part of the project initiatives

millstone
Unit 1 Nuclear Power Station

Strategic Recovery Plan

Short Term

- o Project team put in place
- o Remediation plans
- o Train on 4 common JPMs
- o Enhance system knowledge on generic weaknesses (first cycle 93)
- o Use of advanced classroom training (first cycle, 1993)

millstone
Unit 1 Nuclear Power Station

L.F. SILLIN, JR. NUCLEAR TRAINING CENTER
NU
EXCELLENCE IN OPERATIONS

Strategic Recovery Plan

Short Term

- o Reinforced in plant knowledge with JPM training (first cycle 93)
- o Use of confidence weighted testing results to design future cycles
- o Train all operators on SOER 92-1 - Self checking (first cycle 93)

millstone
Unit 1 Nuclear Power Station

Strategic Recovery Plan

Long Term

- o Project team to remain in place through the recovery effort
- o Exam materials upgraded in accordance with NUREG 1021 requirements
- o Training schedule will incorporate systems training based on revised Job Analysis information
- o JPMs will be structured in program over two year cycle based on revised Job Analysis information

millstone
Unit 1 Nuclear Power Station

Strategic Recovery Plan

Long Term

- o Simulator exercise guides will include additional instructional strategies and standards
- o Periodic independent review of LORT program
- o Assess all operators in third quarter of 1993 to determine effectiveness of pilot effort
- o EOP flow charts trained on prior to July 1, 1993

Exams and Use as Preservatives of Course-Acquired Knowledge

THOMAS K. LANDAUER
KATHLEEN I. AINSLIE
Bell Laboratories

ABSTRACT

Retention of material learned in college-style technical courses was measured by a repeated final examination given one year later. Matched groups of students also repeated the final examination six weeks or six months after the end of the courses. The six-month group showed almost no loss over the year. The six-week group suffered about the same considerable loss as did a control group with no intervening examination. Questionnaire measures indicated that students who used over 10 percent of the examined material in their average monthly work also forgot very little. Apparently, course-acquired knowledge can be kept fresh over long periods by small amounts of properly spaced review activities.

THERE IS WIDESPREAD concern that knowledge and skill acquired through college-style courses are too often lost before they can effectively be put to use by the student. This is a report of an investigation of the use and retention of knowledge from some advanced technical courses over a period of one year, with particular attention to the possible loss-preventing function of retaking examinations after a course is over.

The initial motive for the investigation was to see whether a simple, easy-to-use technique could be developed that would help students to retain more of their new knowledge over long periods. A substantial body of literature in educational and learning psychology demonstrates that tests, recitation, and active practice in general, are effective in reducing the rate of decay of knowledge and skill (*e.g.*, 1, 2, 3, 4, 6, 7, 8). However, the effects of tests on retention over periods of practical significance—months and years rather than days or weeks—do not seem to have been studied. Likewise, tests or other review activities at delays of months have not been studied. College-type courses are often aimed at applications at least a year after their completion. The purpose of this study was to attack directly such very long-term retention and methods of improving it under practical constraints. The proposed technique for providing active practice, and at the same time measuring the retention of course-engendered knowledge over long periods, was as follows. Students took the usual end-of-course examination. Then, at a later time, they were given an opportunity to take the identical examination again, as a "refresher."

In order to measure the long-term effectiveness of the technique, all students were asked to take a terminal examination—again identical—a year after the course was over. In addition, a questionnaire was administered to estimate how much each student used material from his course.

The courses studied were part of the Bell Laboratories In-Hours Continuing Education program. This is a large-scale program offering post-graduate courses in a variety of fields of science, mathematics and engineering. The courses are taught in standard college-style formats in a two-semester schedule. Classes meet about two hours per week for 16 weeks during ordinary work hours, and students are given homework assignments intended to require about four hours per week. Instructors are Ph.D.-level research scientists and engineers. The courses have final exams, and while there is no formal grading, successful completion of a course is recorded in the employee's permanent record. The students are all technical employees of Bell Laboratories. There is a wide range of educational and employment history and ages among the students. Almost all students have at least bachelors degrees, and many hold masters degrees or doctorates.

Prior to the spring semester of 1971, instructors were recruited for participation in the study. Five instructors volunteered to do the considerable extra administration and examination scoring required. Students in their courses were asked to participate on a voluntary basis. Participation required that students take, in addition to the ordinary end-of-course examination, another examination a year later and, in most cases, still another exam sometime between. The students were informed as to the nature of the experiment. However, they did not know whether they would be asked to take the interim exam, or whether they would be in the control group that faced only an exam a year later. They did not know when interim exams would be given. Approximately one-third of the students in these five courses volunteered to participate, and returned in a timely fashion all of the examinations which they were asked to take. The main data from the experiment upon which results are to be reported were obtained from 55 students.

The design of this study allows the following questions to be explored. First, how do the knowledge and skills acquired in such courses last over the succeeding year? Second, how much use is made of the information gained from such a course over the next year? Third, is the material better retained if it is used, as is commonly assumed, and if so, to what extent? Fourth, if students take additional exams after the end of the course, does this aid retention? If so, does it matter when the exams are taken?

Table 1.–Number of students assigned (total) and who returned complete (good) data.

Course	N Students	6 Weeks Group Total	6 Weeks Group Good Data	6 Months Group Total	6 Months Group Good Data	Control Group Total	Control Group Good Data	Total Good Data
Random Processes	4	1	0	1	1	1	1	2
Physical Design	16	5	2	5	0	5	1	3
Digital System Design	11	3	2	3	3	3	3	8
PL-1 Programming Language	32	10	5	10	5	10	5	15
Assembly Language	82	27	5	27	11	27	11	27
TOTALS	145	46	14	46	20	46	21	55

This paper will present results relevant to all of these problems, along with data analyses by which some potential biases arising from the less-than-perfect participation are dealt with. Finally, it discusses what conclusions may reasonably be based on the results.

Design

Titles of the five courses involved, along with the numbers of students who completed various parts of the study, are shown in Table 1. A total of 145 students took examinations at the end of the various courses. These examinations were scored for each class by its instructor, and members of each class were ranked in terms of their scores. The students in each class were then divided into matched triplets (the three best students, the next three, etc.) on the basis of these ranks. From each triplet, one student was picked randomly to be in each of three experimental groups. Where the number in the class was not evenly divisible by three, the extra students were not assigned to experimental groups. The three experimental groups were: a control group that received no examination until one year later; a six-week group that received an examination six weeks after the end of the course and another after a year; and a six-month group that received an examination six months after the end of the course and again after a year.

Table 2 summarizes the design. The end-of-course exam is referred to as the Initial examination, the examinations occurring at six weeks or six months will be referred to as Interim examinations, and the examination occurring after a year will be called the Terminal examination. Note that all students received an Initial examination in June 1971. All students also received a Terminal examination in July 1972. Those in the six-week and six-month groups received Interim examinations in July 1971 or January 1972, respectively.

Methods and Procedures

The examinations were sent by mail to each student at the appropriate time. The examinations, which were Xerox copies of the original exams taken by the students, were accompanied by a cover letter requesting cooperation and instructing the student to take the examination in the same manner as had been required at the end of the course (all courses used take-home exams). The students were assured that their results would be kept anonymous, and they were asked not to put their names on the examination sheets. A code number on each examination was used to keep track of the students, but identities were not available to the instructors who later scored them.

Approximately two weeks after the examinations, follow-up letters were sent to those who had not yet returned them. The number of examinations returned by members of the various experimental groups and classes is shown in Table 1. The proportion of exams returned was about the same for the control and six-month groups, and somewhat lower for the six-week group. The total number of complete sets of data was too small to allow separate analyses by courses; in order to obtain interpretable data, results from all of the courses were combined. Thus it is difficult to characterize precisely the kind of course to which the results apply. However, since well over half of the students were in programming courses, and the rest were in technical or mathematical courses, there is a certain common character to the material.

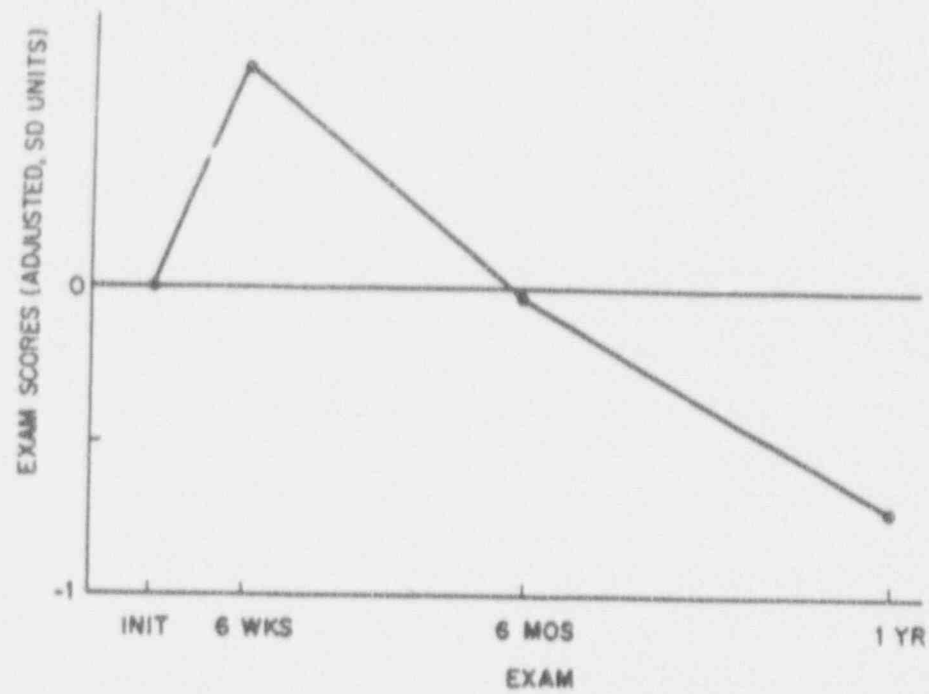
All of the Interim and Terminal examinations were collected before any of them was scored. After they were all returned, in August 1972, the examinations were returned to the original instructors for scoring. Instructors did not know which student had provided which exam paper, nor from which of the examination periods any exam had come. The instructors were asked to review their scoring methods from the previous year and to follow them as closely as possible. They were asked to score all of the examinations together and to try to maintain consistent criteria.

Scores from the Initial examination alone were used to establish a common scale. The Initial scores for each class were first transformed to z-scores. This allowed combining of scores across classes. Raw scores for the Interim and Terminal examinations were converted to the common scale by subtracting the class mean for the Initial examination, and then dividing by the standard deviation of the Initial exam scores. For purposes of establishing the scale, all

Table 2.–Design of Retention Study

	Initial Course Final Exam (June '71)	Interim Exam (July '71)	Interim Exam (January '72)	Terminal Exam (July '72)
Control Group	X			X
6-Week Group	X	X		X
6-Month Group	X		X	X

Figure 1.–Average exam scores over a one year period following technical courses. All scores are based on standard (SD) units derived from original end-of-course final exams (INIT), and are adjusted to equal starting levels. The averages are for independent groups of students who took a given exam first at the end of the course, and next after the time interval indicated. Data are from five Bell Laboratories Continuing Education courses given in spring 1971.



scores for the Initial examinations were used, including those of students who failed to take one or more of the succeeding exams. This was done in order to provide as stable and meaningful a scale as possible.

One question was whether the three experimental groups were similar at the outset. The Initial exam scores for students with complete data in the six-week, six-month and control group were .37, .03 and −.06, respectively. These means do not differ from each other significantly ($F_{2,52} = 0.91$). However, for further analysis, an additional adjustment was made in the scores so as to give all groups equal starting levels. From each of the Interim and Terminal z-scores was subtracted the Initial z-score for the same student.

Several weeks after the Terminal examination, a questionnaire was sent out in order to obtain information about the amount of use that had been made of the material over the preceding year. The questionnaire asked the students to estimate for each month between the end of the course and the Terminal examination the percentage of the material covered on the end-of-course examination that had been used at least once during that month. A calendar was provided in which they were to write their estimated percentages. Eighty-seven questionnaires were returned.

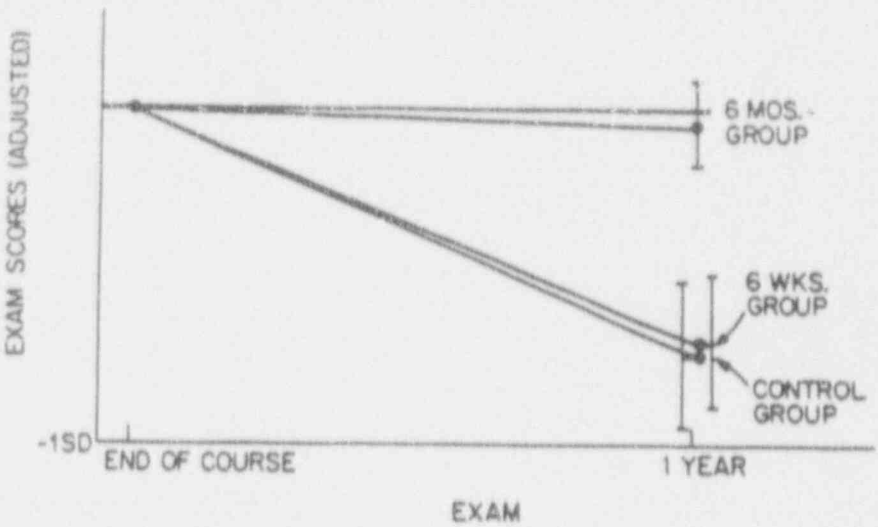
Results

Figure 1 shows the mean adjusted exam scores for the six-week exam, the six-month exam, and the Terminal exam. Note that the Terminal exam score mean is for the control group only. Thus, this figure reflects change in knowledge over a year in the absence of intervention. As can be seen, the mean scores first go up and then decline over the year: 0.01, 0.72, −.03, −.73. At six weeks the scores were significantly higher ($t_{12} = 3.03$, $p < .01$) than they were on the Initial exam, probably because the students found answers to questions they had not known. However, by six months the average student's scores were at the same level or slightly less than they were on the Initial exam. By a year the average student had lost a significant .73 of a standard-deviation-unit of what he had originally learned ($t_{20} = 2.35$, $p < .03$). Some additional intuitive meaning can be gained by considering what such a loss meant for students in a typical course. For example, in one of the courses there was an exam grading scale of 200 points, for which the initial mean was 117 and the standard deviation 30.8. The range of scores was from 44 to 178. Thus, an average student would have slipped from an initial raw score of 117 to 94 by a year later. Stated differently, the loss over a year took the average student from about the 50th percentile on the Initial exam to a score corresponding approximately to the 23rd percentile on that exam.

Considering the effect of the Interim examination, Figure 2 shows the losses over the whole year for students in the control group, the six-week Interim exam group, and the six-month interim exam group. Losses for the six-week and control groups were almost identical; the loss for the six-month group was very little and significantly less than that of the control group ($t_{39} = 1.86$, $p < .05$, one-tailed).

That the effect of a test is greater if the test is delayed

Figure 2.–Decline in adjusted exam scores over a one year period for independent groups having no intervening exam (control) or having an intervening exam after either six weeks or six months. (Vertical lines indicate standard error of means.)



rather than given immediately has precedent in experiments on paired-associated learning in which a non-monotonic effect of delay of testing has been observed (2, 3). Whether the same explanation will serve both for one-minute delays of tests on paired associates as studied by Landauer (2, 3), and for six-month delays of course tests is an open question. If the explanation is that tests as studied here do most good when the S has forgotten some, but not all of what he once knew, then the two phenomena may well be related.

Figure 3 shows the estimated average amount of the material covered on exams which Ss used in various months after the end of the courses. The low range of these estimates seems somewhat surprising. Almost half of the students asserted that they have never used any of the material covered on the examinations. We doubt that this is because the examinations did not cover the important elements of the course, since both the instructors and the investigators were convinced that the examinations used in the courses were quite comprehensive and reflected a good job of choosing representative and important material.

The students who had both taken Terminal examinations and completed questionnaires were divided into three roughly equal-sized groups on the basis of the amount of use they estimated they had made of the course material, averaged over the whole year. The three groups were: (1) 21 who reported that they had never used any of the material; (2) 15 who estimated they used an average of less than 10 percent–for whom the median use per month was 1.7 percent– and (3) 13 who estimated they used 10 percent or more in the average month–for whom the median use per month was 20.0 percent. The adjusted Terminal examination scores for these three groups are shown in Figure 4. Retention was better among those who reported using the material most. The differences shown in Figure 4 are statistically significant ($F_{2,46} = 4.22$, $p < .05$).

The three sets of students differing on estimated usage did not differ systematically or significantly in their Initial test scores, having means of –.05, .43, and .09, respectively, for the none, less-than-10 percent and more-than-10 percent groups ($F_{2,46} = 0.91$). Thus, differences in initial

Figure 3.–Subjective estimates of proportion of material tested on end-of-course final exams used in each of 12 succeeding months.

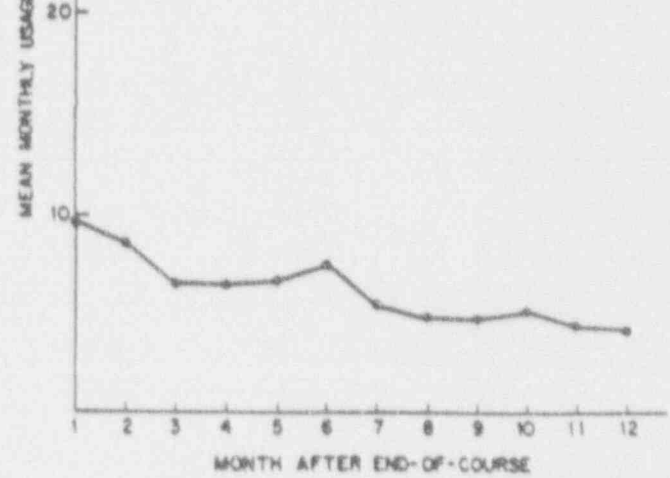
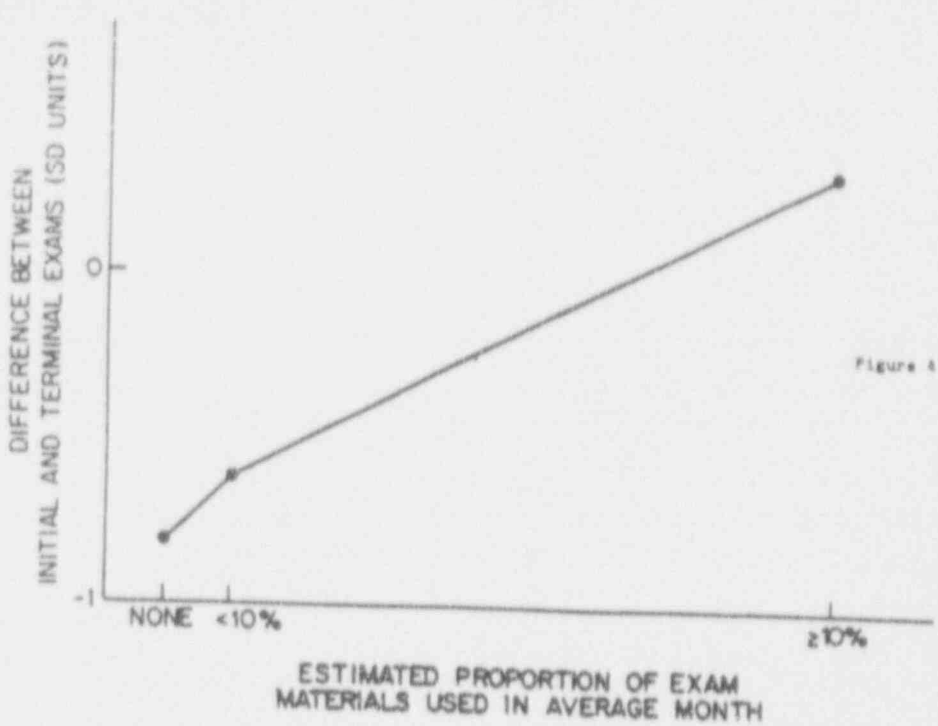


Figure 4.–Change in exam scores over one year period as a function of students' estimates of the amount of examined material which they used in an average month.



talent do not explain the differences in adjusted Terminal scores as a function of the amount of use. This suggests that it was the use that kept the material fresh, rather than that good students used the material more.

Potential Bias Problems

As noted above, the three experimental groups were similar with respect to their scores on the Initial exam. In particular, the control group and the six-month group—the groups whose retention differed significantly over a year—were almost identical on Initial scores. It therefore appears that these groups were similar with regard to original talent. Nonetheless, it is possible that the taking of exams at six weeks or six months served to selectively dissuade students who forgot more from taking the examinations a year later, thus introducing a bias. There were, in fact, 17 students in the six-week and nine in the original six-month group who took the Interim examination but failed to return Terminal examination papers. (One student took the Terminal but not the Interim exam.) We calculated the mean (adjusted) performance of these groups on the Interim examination and they were, indeed, considerably lower than those for the average students in these groups: .43 and –.77 for the six-week and six-month groups, respectively. However, it should be noted that this self-selection is not necessarily indicative of a bias in the comparison of mean results for the experimental groups with the control group. The proportion of students in the six-month group who took the Terminal examination was almost identical to that in the control group; *i.e.*, there was the same overall dropout rate among the control group as among the six-month group. There is no reason to believe that the ones who dropped out of the control group would have done better than those who dropped out of the other groups. Therefore, there is no direct evidence on which to believe that those who remained to take the Terminal examination differed systematically between the groups.

Nonetheless, the investigators felt that it was important to estimate whether the results might have been different had all the students who took the six-week or six-month exams subsequently also taken Terminal exams. This was accomplished by deriving a prediction equation for Terminal scores on the basis of Initial and Interim scores for those students who took all three exams. For example, for the six-month group, a linear regression equation, based on a least-squares fit, showed that Terminal examination scores were predicted as $w = -.03 + .52x + .30y$, where w is the unadjusted Terminal examination z-score, x the unadjusted Initial z-score, and y the unadjusted Interim z-score.

These prediction equations having been determined for the students who took all three exams, they were then applied to the "nontakers," who had taken Initial and Interim, but not Terminal exams. Predicted adjusted scores for them—what they probably would have gotten on the Terminal examinations had they taken them, minus their actual scores on the Initial examinations—were –1.07 for the six-week group and –.36 for the six-month group. The predicted six-week group score was below the actual control group score while the predicted six-month group score was above the control group. For the control group, there is, of course, no corresponding estimate for dropouts. However, on the assumption that control group members

who failed to take the Terminal exam would have done, if anything, somewhat worse than those who actually took it, as predicted for experimental groups, the "true" Terminal score for control students would be even lower than observed. Thus, had all students completed the Terminal exams, it appears very unlikely that the overall direction and significance of the differences in average performance among the three groups would have been changed.

Discussion

The first thing that is worthy of note is the very small amount of use that students claimed to make of the material learned in these courses. The average student seems to have taken a course at least a year before he actually needed it. One might guess that technical courses given in a technical work environment would suffer comparatively little from this problem. The fact that there is a relation between the amount of use and the amount of loss over a year would seem to suggest that it is a potentially serious matter for a student to take a course more than a year before he will use its contents.

On the positive side, an Interim test at six months—but not one at six weeks—seems to help students retain information over a period of a year. Thus, one possible partial solution to the problem of getting knowledge through to its temporal destination is to provide occasional retests, or perhaps other "refresher" activities, for those who continue to anticipate future use. The results of this study suggest that such activities should be spaced about six months apart. While the present study gives evidence of effectiveness only for exams and reported actual use, other active recall exercises such as papers or problem sets should also work. However, examinations, ideally with shortly following corrections, serve the function in an easily administered fashion.

REFERENCES

1. Ausubel, D. P., "Early versus Delayed Review in Meaningful Learning," *Psychology in the Schools*, 3: 195-198, 1966.
2. Landauer, T. K., "Reinforcement as Consolidation," *Psychological Review*, 76: 82-96, 1969.
3. Landauer, T. K.; Eldridge, L., "Effect of Tests without Feedback and Presentation-Test Interval in Paired-Associate Learning," *Journal of Experimental Psychology*, 75: 290-298, 1967.
4. Peterson, H. A.; Ellis, M.; Toohill, N.; Kloess, P., "Some Measurements of the Effects of Reviews," *Journal of Educational Psychology*, 26: 65-72, 1935.
5. Roderick, M.; Anderson, R. C., "Programmed Instruction to Psychology versus Text-book Style Summary of the Same Lesson," *Journal of Educational Psychology*, 59: 381-387, 1968.
6. Rothkopf, E. Z., "Learning from Written Instructive Materials: An Exploration of the Control of Inspection Behavior by Test-like Events," *American Educational Research Journal*, 3: 241-249, 1966.
7. Sones, A. M.; Stroud, J. B., "Review, with Special Reference to Temporal Position," *Journal of Educational Psychology*, 31: 665-676, 1940.
8. Spitzer, H. F., "Studies in Retention," *Journal of Educational Psychology*, 30: 641-657, 1939.

COMPARISON OF THE STANFORD-BINET SCALE WITH THE PEABODY PICTURE VOCABULARY TEST

S. Josephina Concannon
Worcester State College

THE WELL KNOWN TESTS of assessment, the Revised Stanford-Binet Scale of Intelligence and the Wechsler Intelligence Scale for Children, are the commonly used diagnostic tools given individually. Another test less time consuming than these and with a minimum of preparation necessary is the Peabody Picture Vocabulary Test (PPVP).

The test is composed of 150 numbered plates, each containing four pictures; the testee must identify one of the four from a question and/or statement by the tester. Basal and ceiling ages are easily obtained. From the forms (A-B), percentiles and intelligence quotients are obtained for Ss ranging from C.A. of 2-9 to 18-5.

Reliability findings showed coefficients of .97 to .61 based on data from both forms administered to handicapped Ss—deaf, cerebral palsied and mental deviates. Of 11 reliability studies, nine were carried out with a deviate population. The manual states that "coefficients of equivalence and temporal stability appear to be satisfactory for both average children and for those who have one of a number of disabilities." (PPVP manual, p. 32). It is questionable as to what the disability is and how many average children formed part of the standardization population.

No research data on preschool children and on those with above average ability were available. Of the validity studies, 36 percent treated mental retardates, 18 percent examined average children and 46 percent treated various other forms of deviation. Buros cites the spurious effects of the validity coefficients because of the range in age and variance in deviation (1:530).

Table 1.—Mean, Sigma and Standard Error for Stanford-Binet Scale of Intelligence and the Peabody Picture Vocabulary Test

Age	Stanford-Binet Scale of Intelligence				Peabody Picture Vocabulary Test		
	N	M	SD	SE	M	SD	SE
5	32	134.40	6.53	1.16	117.00	14.50	2.58
4	30	130.50	9.35	1.73	114.50	10.10	1.87
3	32	129.30	13.25	2.36	117.75	8.00	1.42

The Spacing Effect

A Case Study in the Failure to Apply the Results of Psychological Research

Frank N. Dempster   *University of Nevada, Las Vegas*

ABSTRACT: The spacing effect would appear to have considerable potential for improving classroom learning, yet there is no evidence of its widespread application. I consider nine possible impediments to the implementation of research findings in the classroom in an effort to determine which, if any, apply to the spacing effect. I conclude that the apparent absence of systematic application may be due, in part, to the ahistorical character of research on the spacing effect and certain gaps in our understanding of both the spacing effect and classroom practice. However, because none of these concerns seems especially discouraging, and in view of what we do know about the spacing effect, classroom application is recommended.

The spacing effect—which refers to the finding that for a given amount of study time, spaced presentations yield substantially better learning than do massed presentations—is one of the most remarkable phenomena to emerge from laboratory research on learning. It is remarkable in several respects. First, the spacing effect is one of the most dependable and replicable phenomena in experimental psychology. Second, it is remarkably robust. In many cases, two spaced presentations are about twice as effective as two massed presentations (e.g., Hintzman, 1974; Melton, 1970), and the difference between them increases as the frequency of repetition increases (Underwood, 1970). Moreover, demonstrations of achievement following massed presentations often are only slightly higher than that following a single presentation (e.g., Melton, 1970). Third, the spacing effect is truly ubiquitous in scope. It has been observed in virtually every standard experimental learning paradigm, with all sorts of traditional research material (Dempster, 1987a; Hintzman, 1974; Melton, 1970).

With all of these characteristics in its favor, the spacing effect would seem to have considerable potential for improving classroom learning. However, there is little evidence that this potential has been realized. Neither American classrooms nor American textbooks appear to implement spaced reviews in any systematic way, and by comparison, Soviet mathematics textbooks provide a much more distributed method of presentation than do their American counterparts (Stigler, Fuson, Ham, & Kim, 1986). Nor is there much evidence that the next generation of educators is being better informed. In a recent sampling of practitioner-oriented textbooks suitable for use in teacher education programs, I found either little or no mention of the practical benefits of the spacing effect, and in some cases the spacing effect was confused with other phenomena (e.g., Good & Brophy, 1986; Mayer, 1987; Slavin, 1986; Woolfolk, 1987). One well-known educator, in fact, advised against spaced practice at least in the early stages of learning (Hunter, 1983).

Why is it that research findings that appear to have significant implications, such as the spacing effect, often are not utilized by teachers and curriculum makers? In general, the problem is that there is no well-developed implementation model, nor is there a standard methodology for analyzing the conditions that foster the transfer of knowledge from the laboratory to the classroom (see Hosford, 1984, for a discussion). Obviously, issues regarding the utilization of findings from basic research are complicated, and there are many potential impediments to the implementation of research findings in the classroom. In this article, I explore nine potential impediments, all of which seem reasonable at first glance, in an effort to determine which, if any, apply to the spacing effect.

Impediments to Application

The Phenomenon Has Not Been Known Long Enough

Although the time lag between discovery and application varies greatly, some considerable period of time often intervenes between the publication of research findings and their application. In the case of the spacing effect, however, a considerable period of time already has passed since its initial documentation. The spacing effect was known as early as 1885 when Ebbinghaus published the results of his seminal experimental work on memory. With himself as the subject, Ebbinghaus found that for a single 12-syllable series, 68 immediately successive repetitions had the effect of making possible an errorless recital after seven additional repetitions on the following day. However, the same effect was achieved by only 38 distributed repetitions spread over three days. On the basis of this and other related findings, Ebbinghaus concluded that "with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time" (Ebbinghaus, 1885/1913, p. 89). Jost, also working with non-

Vol. 43, No. 8, 627-634

sense syllables, reported similar findings and in 1897 formulated what was to become known as Jost's Law: "If two associations are of equal strength but of different age, a new repetition has a greater value for the older one" (McGeoch, 1943, p. 140).

In 1928, Ruch published a review of dozens of studies of the spacing effect. Although interpretation of the results of these studies (e.g., Dearborn, 1910; Perkins, 1914; Pyle, 1913; Starch, 1912) is complicated by other, potentially confounded variables, the results tend, in general, to confirm the earlier work by Ebbinghaus and by Jost. Thus, published reports of the spacing effect have been in existence since the latter part of the 19th century and the early part of the 20th century.

The Phenomenon Has Not Received Recent Documentation

In the absence of recent documentation, research findings may seem stale or anachronistic, but, as most, if not all, students of the learning literature know, the spacing effect has been well-documented in recent times. Many studies of this phenomenon were published during the 1960s and the 1970s, as reviews by Hintzman (1974), Melton (1970), and Glenberg (1979) attest.

Although much of the research included in these reviews was reminiscent of the work of Ebbinghaus in using easily analyzable simple verbal units, the fruits of this research are considerable from any perspective. For example, the ubiquitous, highly replicable character of the spacing effect fostered the notion that its existence must be telling us something important about memory (e.g., Hintzman, 1974). Also, it clearly demonstrated that the Total Time Law, which states that the amount learned is a direct function of study time regardless of how that time is distributed, was in deep trouble or at least in need of a major overhaul (Melton, 1970; Underwood, 1970).

More recently, the spacing of repetitions has been the subject of studies reported in a variety of journals, including some with an applied perspective (Bahrick & Phelps, 1987; Cuddy & Jacoby, 1982; Dellarosa & Bourne, 1985; Dempster, 1987b; Elmes, Dye, & Herdelin, 1983; Glenberg & Lehmann, 1980; Glover & Corkill, 1987; Toppino & DiGeorge, 1984; Toppino & Gracen, 1985). Thus, documented evidence of the spacing effect has appeared in the literature continually for the past 100 years.

The Phenomenon Cannot Be Linked to Issues of Current Concern to Educators

I agree with Glaser (1982) that research knowledge is most likely to inform educational practice if it can be related explicitly to large-scale educational issues or macro variables. Whereas the relationships between the rather fine-grained analyses of learning and memory conducted by researchers and the practice of education usually are not self-evident to the teacher and must be illuminated by the psychologist, this task should be relatively straightforward in the case of the spacing effect. After all, the spacing effect has immediate and obvious implications for how time in the classroom may be distributed optimally. In the wake of recent critiques and studies of schooling, such as *A Nation at Risk* (National Commission on Excellence in Education, 1983), *Time to Learn* (Denham & Lieberman, 1980), and *Perspectives on Instructional Time* (Fisher & Berliner, 1985), the use of time in the classroom has become a major educational concern.

The Phenomenon Has Not Been Demonstrated Satisfactorily in School-Like Activities

Such demonstrations are arguably the most important bridge between basic research and educational practice. In the case of the spacing effect, however, this bridge seems to have been crossed. Several demonstrations of the spacing effect reviewed by Ruch (1928) were, as he put it, "intended for schoolroom application" (p. 20). One of the most interesting of these effects, from an educational perspective, is that of Pyle (1913), who had a group of third-graders drilled in addition, either twice a day for 5 days (once in the morning and once in the afternoon) or once a day for 10 days. Their improvement in recall of addition facts, which was decidedly in favor of the latter instructional method, provided perhaps the earliest experimental confirmation of William James' (1901) advice to teachers and students that it is better to repeat an association on many different days than again and again on just a few days (p. 129).

Most of the early demonstrations of the spacing effect "intended for schoolroom application" focused on text processing tasks, and in three recent studies text processing again has been the focus. In a study by Kraft and Jenkins (1981), subjects attempted to free-recall the title and one idea unit from each of a series of twice-presented passages, with each repetition separated by lags of up to eight intervening passages. As in much standard verbal learning research, recall was a linear and much improved function of lag. One practical limitation of this study, of course, is that students rarely are asked to recall so little from a series of passages.

In the second of these recent spacing effect studies, Dempster (1986) contrasted lags—defined as the interval between two opportunities to read a passage of text material—of 30 seconds, 5 minutes, 20 minutes, and 48 hours in one experiment, and 5 minutes and 30 minutes in a second experiment. In the first experiment, subjects in the 48-hour condition recalled significantly more idea units than did subjects in either of the two shortest lag conditions. In the second experiment, recall was significantly higher in the 30-minute condition than in the 5-minute condition. In both experiments, the recall advantage associated with the best performing group was about

Preparation of this article was supported, in part, by a grant from the Barrick Research Council, University of Nevada, Las Vegas. I thank my colleagues John A. Bates and Doris Carey for their assistance, as well as Arthur Glenberg, Douglas Hintzman, and an anonymous reviewer for their helpful reactions to an earlier version of this article.

Correspondence concerning this article should be addressed to Frank N. Dempster, Department of Educational Psychology, University of Nevada, Las Vegas, NV 89154.

the same (between 25% and 30%), even though different text passages were used in the two experiments. Thus, the effect was both robust and replicable. Finally, Glover and Corkill (1987) observed the spacing effect (0 lag versus a 30-minute lag) in subjects' memory for paragraphs they read as well as for brief lectures (125 words) they heard.

In addition to text recall, spacing effects have been demonstrated in programmed instruction, where the objective has been the learning of science and mathematical concepts. In one study, the meanings of a series of programmed scientific terms were learned much more effectively when repetitions were spaced than when they were massed (Reynolds & Glaser, 1964). In another study, arithmetical rules presented by a computer-assisted instruction system and expressed as verbal statements were learned better when reviews occurred one and seven days after original learning than when they occurred one and two days after original learning (Gay, 1973, Experiment 2).

Finally, spacing effects have been found in vocabulary learning. In a study by Dempster (1987b), 38 uncommon English words and their definitions were presented three times, either with each repetition of any given word separated by every other word (i.e., each repetition of a word was separated by 37 other words or 4 minutes, 19 seconds) or with each repetition of a word massed in succession. In addition, the words were presented either with or without sentence contexts. The results were quite clear. In three experiments in which spaced versus massed presentations were evaluated in this manner, spaced presentations yielded substantially higher levels of vocabulary learning than did massed presentations. In some cases, in fact, the number of word meanings recalled was over 50% greater under spaced conditions than under massed conditions.

In Dempster's (1987b) study the retention interval averaged less than an hour—short from a practical perspective. Thus, the question might be asked, "Would spacing effects emerge in the retention of vocabulary words tested at much longer intervals?" Fortunately, there are data relevant to this question. Bahrick and Phelps (1987) tested 35 individuals who had learned and relearned 50 English–Spanish word pairs for recall and recognition after an interval of eight years. One variable of interest was the interval between successive relearning sessions—either 30 days, 1 day, or 0. The data show that the intersession interval had a very large effect on recall, with the recall probability associated with the 30-day interval about 2.5 times the probability associated with the zero interval. In turn, the 1-day interval was associated with much better retention than the zero interval. For both comparisons, the effect on recognition, exclusive of recall, was much less pronounced. Moreover, their data clearly indicate that even five or more presentations are unlikely to facilitate long-term retention if the interval between successive presentations is one day or less. With respect to the educational implications of their study, the authors concluded that long-term retention would almost certainly be enhanced if foreign language courses

> make certain that students independently retrieve target information at intervals that are as long as 1 month, over a period of several years, instead of the more typical intervals of 1 to 2 days over periods of from 10 to 15 weeks. (p. 349)

There Are Serious Discontinuities in the Literature on the Spacing Effect

Another possible reason for the failure to apply the results of research on the spacing effect is that there are serious discontinuities in the literature on the spacing effect, such that most recent studies seem uninformed by the research of earlier ones. Although the spacing effect has a lengthy published history, there are discontinuities of this sort. For example, much of the important work from an applied perspective done in the early 1900s is not cited by studies published later. More generally, most recent studies tend to give the impression that, with the exception of the work of Ebbinghaus and Jost, all we know about the spacing effect dates only as far back as the 1960s. I cannot help but wonder how widespread a problem this is in the learning literature, and what sorts of consequences it has for the evolution of the science of learning and its application. Why is it that we occasionally—and perhaps frequently—give up on, or simply lose interest in, a phenomenon before we have definitive answers to basic questions and, then much later, return to the phenomenon as though we had just recently discovered it?

Upon reflection, this ahistorical character of research on the spacing effect would seem to have at least two unfortunate consequences, either of which could impede application. From a scientific perspective, such research is less likely to be as cumulative—where cumulative refers to empirical laws and theoretical structures building on one another so that later developments extend and unify earlier work (Hedges, 1987)—as would be more historical research. From a practical perspective, it will seem as if the spacing effect has not weathered continuous scrutiny over a lengthy period of time.

There are, of course, many possible reasons for discontinuities in the literature of a phenomenon. In the case of the spacing effect, however, three appear to stand out. First, for the most part, studies conducted from an applied perspective and those conducted from a basic research perspective constitute two distinct streams of research. For example, although widely cited reviews by Crowder (1976), Glenberg (1979), Hintzman (1974), and Melton (1970) report dozens of studies of the spacing effect using simple list learning materials, none of the many studies using more complex material with clearer classroom analogues are reported. This is the case despite the fact that some still viable accounts of the spacing effect (see for example, Cuddy & Jacoby, 1982, and Hintzman, 1974) were anticipated and supported in earlier studies done from a more applied perspective (see, for example, Ausubel, 1966).

Second, different terminology has been used to refer to similar, though distinguishable, phenomena—a situation that seems to have created some confusion. For example, a number of writers have distinguished between

the "spacing effect" and both the "Melton" or "lag effect" and the "massed-versus-distributed practice" effect. Lag effects have been said to occur when performance improves as a function of the number of intervening items between successive presentations, whereas massed versus distributed practice effects often have been restricted to comparisons between spacings of zero (massed practice) and all spacings greater than zero (distributed practice). The use of this terminology, however, is somewhat uneven. Some researchers have used these terms interchangeably or have included other variables (e.g., length of period of study, retention interval) in their characterizations of the spacing effect (e.g., Ruch, 1928). Also, much of the massed versus distributed practice research, which has focused on perceptual motor skills tasks and lists of nonsense syllables, has yielded weak effects of spacing. According to Underwood (1961), who reviewed 10 years of distributed practice research, "Even under the most favorable conditions for facilitation by distributed practice, one could not recommend its use in an applied setting" (p. 230). Apparently following his lead, some older educational psychology texts advised that there were no clear practical implications to be drawn from distributed practice research (e.g., DeCecco, 1968; Mathis, Cotton, & Sechrest, 1970).

Third, the spacing effect is just one of a family of similar, though less thoroughly investigated, phenomena that are occasionally confused in the literature. One such phenomenon is the so-called "test-spacing" effect, which refers to the fact that spaced tests, particularly tests with intertest intervals of an expanding nature, result in greater retention than do massed testings (Landauer & Bjork, 1978; Rea & Modigliani, 1985; Spitzer, 1939). Another related phenomenon has been observed when once-presented written exercises or materials in a short course in statistics either are spread over the course of several sessions or are presented in a single session—a situation that is analogous to "cramming" for a test. In this case, students learn more when the material is distributed over several sessions (Bloom & Shuell, 1981; Smith & Rothkopf, 1984). Finally, Reder and Anderson (1982) found that, with total study time equated, repeated, well-spaced presentations of a text were more effective than was a single, longer presentation. Similarly, Edwards (1917) compared groups who studied various school materials, including history and geography, six and one-half minutes continuously or with the same amount of time divided into a study period of four minutes followed some days later by a review of two and one-half minutes. Lag and test intervals, difficulty of the material, and age of the subjects were so variable that it is difficult to interpret the results; however, without exception they favor the repeated, spaced-study groups.

To complicate matters further, research on allied phenomena also has an ahistorical character. For example, Rea and Modigliani (1985) failed to cite Spitzer (1939), even though Spitzer's work showed that if the interval between original learning and the first test in a series is too lengthy, test spacing effects are likely to be vitiated. Also, Reder and Anderson (1982) failed to cite Edwar (1917), and neither Bloom and Shuell (1981) nor Smi and Rothkopf (1984) seemed aware of the work of A (1950), who found only minor differences in retenti between groups treated like theirs.

Too Many Studies Using School-Like Activities Have Failed to Show the Spacing Effect

Although the spacing effect is one of the most dependal phenomena in the learning of standard verbal learni lists, there have been more than a half-dozen document failures to observe the spacing effect in tasks with cla room analogues. The results of these studies make it qu clear that the spacing effect is subject to certain not fu understood boundary conditions. Specifically, five so of boundary conditions are suggested. First, it has be found that under certain circumstances spaced prese tations are no better than (Austin, 1921) and sometim even worse than (Gordon, 1925) massed presentations tests of immediate recall. For example, Austin found th massed readings (e.g., five times in one day) of text m terial proved as effective as spaced readings (e.g., da for five days) in tests of immediate recall, whereas t spaced readings were much more effective in delayed te particularly if they came two to four weeks after learni Second, it has been found that massed practice often more efficient for certain simple, isolated skills, such writing the products of number pairs as rapidly as possi (Thorndike, 1916). Third, evidence from traditio learning research suggests that the spacing effect may r apply to preschool age children, although it does eme in a robust manner by age seven (Toppino & DiGeor 1984). Fourth, two studies have shown that the spac effect can be eliminated if paraphrased rather than v batim versions of the repeated materials are used (D larosa & Bourne, 1985; Glover & Corkill, 1987).

Finally, the results of a number of studies seem suggest that beyond a certain lag interval, further increa in lag are not always associated with further increases learning. For example, English, Wellborn, and Kill (1934) found that four readings of a text at three-h intervals were associated with better learning than f consecutive unspaced readings; however, readings at thr hour intervals were no better than readings at either o or three-day intervals. Similarly, Lyon (1914), Peters Ellis, Toohill, and Kloess (1935), and Sones and Str (1940) reported essentially no differences in retention tween groups with rereading reviews spaced 1 and 7 and 9, and 1 and 17 days after original learning. Th findings were later corroborated by Ausubel (1966) by Gay (1973).

Of these boundary conditions, the last two seem most serious and most puzzling in view of the fact t the spacing effect has been found in paraphrased mate (Rothkopf & Coke, 1966) and in view of long-lag eff obtained with traditional verbal learning material (G berg & Lehmann, 1980) and in vocabulary lear (Bahrick & Phelps, 1987). With respect to the latte may be that under certain lengthy lag conditions, the u

benefits of spaced repetitions do not obtain because the results of initial processing efforts have been forgotten (see Lyon, 1914, and Sones & Stroud, 1940, for earlier discussions of this hypothesis).

In any case, the spacing effect, especially in text processing, cannot be taken for granted. Although there have been relatively few documented failures to obtain the spacing effect, they certainly could have raised enough doubt about the dependability of the effect to discourage application.

The Phenomenon Has Not Been Demonstrated Satisfactorily in the Classroom

With very few exceptions (Dempster, 1986; Glover & Corkill, 1987; Pyle, 1913), even research using educationally relevant materials has been conducted in the laboratory. Moreover, the rather simple learning situations created in classroom studies of the spacing effect fail to approach the complexities facing curriculum developers and teachers. Curriculum developers and teachers have to concern themselves with the design of instruction conveyed in classes that often meet every day for a school term, under conditions in which much if not most of the content is organized in ways that imply systematic movement through learning hierarchies, curricular sequences, and so forth. Thus, it may be argued that it is not at all clear what specific implications demonstrations of the spacing effect in simple, isolated classroom situations have for curriculum designers and teachers faced with decisions about how much material to include in a course, how to sequence it, and how to optimally phase in new material and phase out old material.

Arguably, the relative lack of applied research in educational settings is, from an educational perspective, the most serious shortcoming of research on the spacing effect. There is no substitute for applied research, and the absence of at least several convincing demonstrations of the spacing effect in ongoing classroom situations under naturalistic conditions may well have been an impediment to application.

Too Little Is Known About Actual Classroom Practice to Justify Widespread Application of the Spacing Effect

To the best of my knowledge, nothing has been published concerning the proportion of time in the classroom teachers usually devote to review (i.e., any re-presentation or practice activity pertaining to a particular educational objective), and to what extent reviews are massed as opposed to spaced. Moreover, the general nature of classroom review activities apparently has not been characterized. For example, to what extent are reviews verbatim or paraphrased? Because several studies have failed to show spacing effects in text processing when paraphrased rather than verbatim versions of the repeated materials were used (Dellarosa & Bourne, 1985; Glover & Corkill, 1987), the answer to this question is of practical importance.

In short, due to significant gaps in our understanding of classroom practice, the application potential of the spacing effect cannot be estimated with any precision. Accordingly, it may be assumed wrongly that efforts to implement the spacing effect would result in little benefit.

The Phenomenon Is Not Sufficiently Understood

Phenomena that are not well understood are likely to invite skepticism among practitioners—particularly those who are familiar with any one of several instances in which the application of a poorly understood finding has had extremely unfavorable consequences (e.g., thalidomide). For example, educators, who often give the impression of having a low regard for fact memorization, might feel that the spacing effect would interfere with the operation of "more laudable, higher mental processes," because it is exactly such memorization to which spacing applies most clearly.

In fact, the theoretical picture surrounding the spacing effect is confused and uncertain, despite numerous attempts at clarification (e.g., Dellarosa & Bourne, 1985; Hintzman, 1974). A recent case in point is a published failure to replicate findings implicating a component-levels interpretation of the spacing effect in standard verbal learning tasks (Toppino & Gracen, 1985). The component-levels hypothesis is one of a group of hypotheses that attribute the spacing effect to increasing independence of encoding events with increasing intervals between repetitions.

With more naturalistic material, theoretical work has been slow to develop, and little has been done in the way of theoretically derived hypothesis testing. However, it is worth noting that Ausubel (1966), in a rarely cited study, offered the following explanation for the advantages of spaced review in meaningful learning:

> In the first place, after a longer retention interval, when more material is forgotten, the learner is more highly motivated to profit from the opportunity for review. He is less likely to regard this opportunity as unnecessary and superfluous and is hence more disposed to take good advantage of it in terms of effort, attention, and concentration. (p. 197)

Apparently, Ausubel (1966) was on the right track when he stressed the role of attention and effort in the spacing effect. Recent evidence indicates that some sort of attentional account of the spacing effect is at least as viable as any other account and well worth pursuing (Dellarosa & Bourne, 1985; Dempster, 1986; Elmes, Dye, & Herdelin, 1983; Magliero, 1983). For example, Dempster (1986) had subjects respond to a questionnaire administered following a recall test of a twice-read passage, with the two readings spaced either 30 minutes apart or 5 minutes apart. The questionnaire consisted of 10 items, each followed by a 10-point rating scale, which was designed to elicit self-reports of various cognitive and affective states and processes during reading and testing. Included were questions concerning levels of attention, interest, anxiety, rehearsal, and changes of interpretation from one reading to the next. The results were quite clear. Significant group differences emerged on only two of the

items, specifically one asking the subjects to indicate how "interested" they were during the second reading—an affective state—and one asking them to indicate how much "attention" they paid during the second reading—a cognitive process. In both cases, the average ratings of students in the spaced 30-minute condition (those who also did best on the recall test) were higher than those in the massed 5-minute condition. Moreover, a correlational analysis, applied to the scores of both groups combined, revealed a significant correlation between recall and only one of the questionnaire items—that is, the attention paid during the second reading. Those who reported having paid more attention tended to have learned more from the text.

However, why should spaced presentations be more interesting (see also Elmes et al., 1983) and receive more attention than massed presentations? Here again, Ausubel (1966) seemed to have anticipated more recent developments when he suggested that it has to do with the relative accessibility of previous encodings (in his words, the activity of "trying and failing to remember material," p. 197). If a student receives massed presentations, the information learned during earlier presentations should be relatively easy to retrieve from memory during subsequent presentations. Thus, subsequent presentations should be relatively redundant or familiar and thus relatively boring. By contrast, if a student receives two well-spaced presentations, the information learned during initial presentations should be relatively inaccessible during subsequent presentations, which should heighten interest level and the amount or quality of attention subsequent presentations receive. According to this account, then, massed presentations are relatively ineffective because they may not actually result in much repetitive processing (see also Cuddy & Jacoby, 1982; Dellarosa & Bourne, 1985; Greeno, 1970; Jacoby, 1978; Underwood, 1970).

One implication of this account is that anything that increases the likelihood that a repetition will receive full processing, such as events that make it difficult to retrieve the results of prior encodings, should improve learning. Thus, this account helps to explain failures to obtain the spacing effect with paraphrased repetitions, that is, repetitions having a changed surface structure (Dellarosa & Bourne, 1985; Glover & Corkill, 1987), and under lengthy lag conditions (Ausubel, 1966; English et al., 1934; Gay, 1973; Lyon, 1914; Peterson et al., 1935; Sones & Stroud, 1940).

In sum, although recent studies have yielded some promising clues to the mechanisms underlying the spacing effect, our theoretical ignorance may have been and may continue to be an impediment to application or might contribute to inappropriate applications.

Summary and Conclusions

The spacing effect would seem to have considerable potential for improving classroom learning, yet there is no evidence of widespread application. In this article, I have considered nine possible impediments to the implementation of research findings in the classroom. Of the nine, five appear to apply to the spacing effect. These include the ahistorical character of research on the spacing effect, some failures to obtain the effect with school-like activities, a paucity of impressive classroom demonstrations of the phenomenon, limited knowledge of classroom practice, and an incomplete understanding of the psychological bases of the spacing effect. By contrast, the fact that the phenomenon (a) has been known for a long period of time, (b) has received recent documentation, (c) can be linked to current educational issues, and (d) has been shown to extrapolate to school-like activities suggests that the first four dimensions of analysis considered do not apply.

The following question now arises: To what extent *should* the list of plausible impediments to the application of the spacing effect in the classroom discourage application? After addressing each of these concerns, I conclude that we do know enough about the effect of spacing to make a very strong argument for application without any additional knowledge about the spacing effect or classroom practice.

Consider first the ahistorical and somewhat confusing character of research on the spacing effect. Given the long history of research on the spacing effect and the considerable recent documentation, the discontinuities in the literature on the spacing effect do not seem terribly important. In fact, the spacing effect has weathered continuous scrutiny over a lengthy period of time. Of course, the fact that the spacing effect is just one of a family of similar phenomena would seem only to strengthen and broaden the appeal of "spacing" in the classroom.

A second plausible reason for the absence of widespread application of the spacing effect is that there have been some failures to obtain the effect in school-like tasks. The spacing effect, in fact, does appear to be subject to certain, not fully understood boundary conditions. However, it would be unrealistic to expect the spacing effect to apply in every situation, and the relatively few failures to obtain the spacing effect seem trivial in light of its many demonstrations. For this reason and in the absence of any serious contraindications to the application of the spacing effect, the fact that the spacing effect does not always work hardly seems to justify resistance to its application.

Arguably, the most serious of the plausible impediments to the application of the spacing effect is the paucity of impressive classroom demonstrations of the phenomenon. Clearly, programmatic research on the effects of spacing in education settings is long overdue, as the results of such efforts would likely aid in its application. The most useful studies of this sort would be those involving curriculum design and classroom teaching that help shed light on the implications of the spacing effect for specific applied issues, such as homework, frequency of testing and feedback, learning hierarchies, mastery learning, and questioning strategies.

Although additional classroom research is desirable and even necessary in order to make the most effective

use of spacing. I do not think we have to withhold all judgment on the applicability of the spacing effect at this point. For example, we know that children study vocabulary in a variety of subjects and that the spacing effect facilitates vocabulary learning. Likewise, children are exposed to numerous scientific terms and arithmetical rules while in school—each of which is a domain in which spacing has been found to be effective. When coupled with what we already know about the effects of spacing in text processing, there is ample reason to believe that the spacing effect will improve classroom learning in a wide variety of subject areas.

Although we do not know much about current classroom applications of the spacing effect, there are many reasons to believe that the spacing effect is underutilized in the classroom in terms of its potential for improving learning. In addition to the reasons indicated earlier, there is the fact that the spacing effect is somewhat counterintuitive. The spacing effect typically refers to a phenomenon that occurs under conditions in which the retention interval between the last presentation and the test is held constant. Thus, one might reason that because the retention interval between the first presentation and the test is shorter under massed conditions, this condition should result in superior performance. Even experienced educators, when judging the instructional effectiveness of text passages, tend to rate prose in which the repetition of a given unit of information is massed as better than those in which it is spaced (Rothkopf, 1963).

In short, the spacing effect is neither intuitively obvious, nor well known among educators. Accordingly, it is reasonable to assume that those who become teachers, administrators, curriculum developers, or writers of reading series are ignorant of the spacing effect, just as many psychologists are not clear about the totality of educational situations that call for its application. Thus, our ignorance of actual classroom practice should not be interpreted to mean that widespread implementation of the spacing effect has little or no potential for improving classroom learning.

The final plausible impediment to the application of the spacing effect is that it is not well understood theoretically. However, in view of the absence of evidence linking the spacing effect to some undesirable psychological process or outcome, this impediment is extremely hard to justify. For example, there is no evidence that the spacing effect impairs the ability to conceptualize or to think critically. Even if it is found that the psychological basis of the spacing effect applies only or largely to memorization tasks, memory is of central importance to any complex intellectual activity.

Nevertheless, a fuller understanding of the spacing effect could eventually aid in its application and might help to avoid some inappropriate applications. Hence, I believe the search for the underlying cause(s) of the spacing effect should proceed. As part of this effort, some attempt should be made to determine if spaced presentations tend simply to increase the amount of information learned (the quantitative hypothesis), or if they tend to increase the learning of only certain kinds of information (the qualitative hypothesis). So far, these hypotheses have been examined only in relation to the more general issue of how repetitions improve learning, with mixed results (Annis & Annis, 1987; Bromage & Mayer, 1986; Mayer, 1983). They have not been examined in the context of comparisons of massed and spaced repetitions, although such comparisons might be fruitful from both a theoretical and a practical perspective.

Recently, I was told that the spacing effect has been "studied to death," and on another occasion that "we know all that we need to know about the spacing effect." Clearly, the spacing effect is one of the most studied phenomena in the 100-year history of learning research, and we *do* know a lot about it—enough to recommend application. However, it would be a mistake to do what these comments imply—which is simply to stop investigating the phenomenon. Although it may take some clever research to avoid diminishing returns, continued experimental study of the spacing effect can yield valuable information regarding its parameters and cause(s). Then too, applied research and widespread application should produce the data base needed to evaluate the particular conditions under which the spacing effect works best.

REFERENCES

Annis, L. F., & Annis, D. B. (1987, April). *Does practice make perfect? The effects of repetition on student learning.* Presented at the annual meeting of the American Educational Research Association, Washington, DC.

Ash, P. (1950). The relative effectiveness of massed versus spaced film presentations. *Journal of Educational Psychology, 41,* 19–30.

Austin, S. D. M. (1921). A study in logical memory. *American Journal of Psychology, 32,* 370–403.

Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools, 3,* 195–198.

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

Bloom, K. F., & Shuell, T. J. (198.). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74,* 245–248.

Bromage, B. K., & Mayer, R. E. (1986). Quantitative and qualitative effects of repetition on learning from technical text. *Journal of Educational Psychology, 78,* 271–278.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior, 21,* 451–467.

Dearborn, W. F. (1910). Experiments in learning. *Journal of Educational Psychology, 1,* 373–388.

DeCecco, J. P. (1968). *The psychology of learning and instruction: Educational psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition, 13,* 529–537.

Dempster, F. N. (1986). *Spacing effects in text recall: An extrapolation from the laboratory to the classroom.* Manuscript submitted for publication.

Dempster, F. N. (1987a). Time and the production of classroom learning: Discerning implications from basic research. *Educational Psychologist, 22,* 1–21.

Dempster, F. N. (1987b). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79,* 162–170.

Denham, C., & Lieberman, A. (1980). *Time to learn.* Washington, DC: National Institute of Education.

Ebbinghaus, H. (1913). *Memory* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College. (Original work published 1885; paperback ed., New York: Dover, 1964)

Edwards, A. S. (1917). The distribution of time in learning small amounts of material. In *Studies in Psychology: Titchener commemorative volume* (pp. 209–213). Worcester, MA: Wilson.

Elmes, D. G., Dye, C. J., & Herdelin, N. J. (1983). What is the role of affect in the spacing effect? *Memory & Cognition, 11,* 144–151.

English, H. B., Wellborn, E. L., & Killian, C. D. (1934). Studies in substance memorization. *Journal of General Psychology, 11,* 233–260.

Fisher, C. W., & Berliner, D. C. (1985). *Perspectives on instructional time.* New York: Longman.

Gay, L. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology, 64,* 171–182.

Glaser, R. (1982). Instructional psychology: Past, present, and future. *American Psychologist, 37,* 292–305.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7,* 95–112.

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8,* 528–538.

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology, 79,* 198–199.

Good, T. L., & Brophy, J. E. (1986). *Educational psychology* (3rd ed.). New York: Longman.

Gordon, K. (1925). Class results with spaced and unspaced memorizing. *Journal of Experimental Psychology, 8,* 337–343.

Greeno, J. G. (1970). Conservation of information-processing capacity in paired-associate memorizing. *Journal of Verbal Learning and Verbal Behavior, 9,* 581–586.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? *American Psychologist, 42,* 443–455.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Potomac, MD: Erlbaum.

Hosford, P. L. (Ed.). (1984). *Using what we know about teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Hunter, M. (1983). *Increasing your teaching effectiveness.* Belmont, CA: The Learning Institute.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667.

James, W. (1901). *Talks to teachers on psychology: And to students on some of life's ideals.* New York: Holt.

Kraft, R. N., & Jenkins, J. J. (1981). The lag effect with aurally presented passages. *Bulletin of the Psychonomic Society, 17,* 132–134.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes *Practical aspects of memory* (pp. 625–632). New York: Academic Press.

Lyon, D. O. (1914). The relation of length of material to time taken for learning and the optimum distribution of time. *Journal of Educational Psychology, 5,* 1–9, 85–91, 155–163.

Magliero, A. (1983). Pupil dilations following pairs of identical and related to-be-remembered words. *Memory & Cognition, 11,* 609–615.

Mathis, B. C., Cotton, J. W., & Sechrest, L. (1970). *Psychological foundations of education.* New York: Academic Press.

Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advanced organizers on learning from science prose. *Journal of Educational Psychology, 75,* 40–49.

Mayer, R. E. (1987). *Educational psychology: A cognitive approach.* Boston: Little, Brown.

McGeoch, J. A. (1943). *The psychology of human learning.* New York: Longmans Green.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596–606.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

Perkins, N. L. (1914). The value of distributed repetitions in rote learning. *British Journal of Psychology, 7,* 253–261.

Peterson, H. A., Ellis, M., Toohill, N., & Kloess, P. (1935). Some measurements of the effects of reviews. *Journal of Educational Psychology, 26,* 65–72.

Pyle, W. H. (1913). Economical learning. *Journal of Educational Psychology, 3,* 148–158.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning, 4,* 11–18.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition, 10,* 97–102.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55,* 297–308.

Rothkopf, E. Z. (1963). Some observations on predicting instructional effectiveness by simple inspection. *Journal of Programmed Instruction, 3,* 19–20.

Rothkopf, E. Z., & Coke, E. V. (1966). Variations in phrasing and repetition interval and the recall of sentence materials. *Journal of Verbal Learning and Verbal Behavior, 5,* 86–91.

Ruch, T. C. (1928). Factors influencing the relative economy of massed and distributed practice in learning. *Psychological Review, 35,* 19–45.

Slavin, R. E. (1986). *Educational psychology: Theory into practice.* Englewood Cliffs, NJ: Prentice-Hall.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1,* 341–358.

Sones, A. M., & Stroud, J. B. (1940). Review with special reference to temporal position. *Journal of Educational Psychology, 31,* 665–676.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641–656.

Starch, D. (1912). Periods of work in learning. *Journal of Educational Psychology, 3,* 209–213.

Stigler, J. W., Fuson, K. C., Ham, M., & Kim, M. S. (1986). An analysis of addition and subtraction word problems in American and Soviet elementary mathematics textbooks. *Cognition and Instruction, 3,* 153–171.

Thorndike, E. L. (1916). Notes on practice, improvability and the curve of work. *American Journal of Psychology, 27,* 550–565.

Toppino, T. C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12,* 118–122.

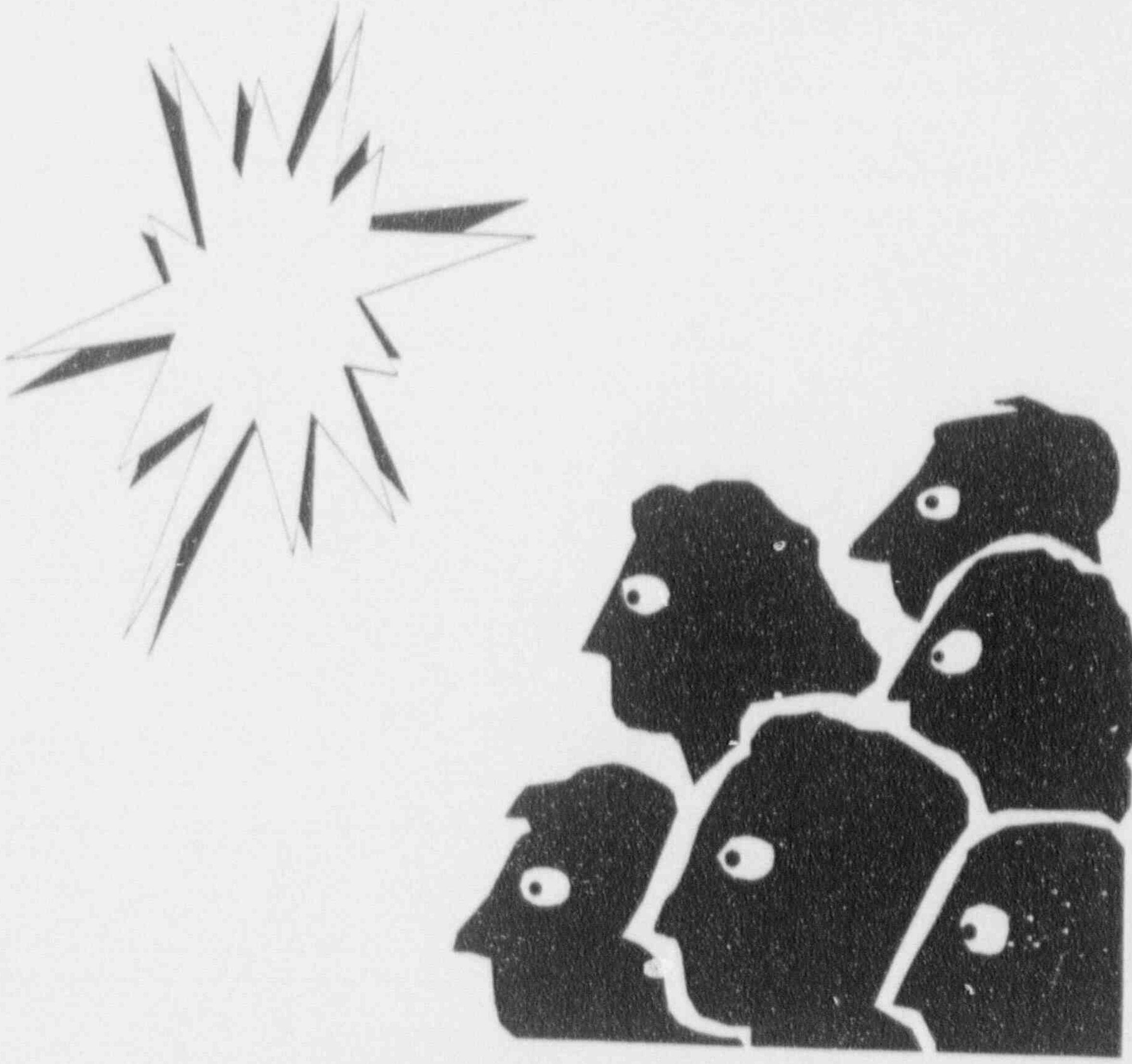
Toppino, T. C., & Gracen, T. F. (1985). The lag effect and differential organization theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 185–191.

Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review, 68,* 229–247.

Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 573–580.

Woolfolk, A. E. (1987). *Educational psychology* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

VISUALIZATION TRAINING PROGRAM



U.S. Olympic Training Center
Colorado Springs, Colorado

IMAGERY TRAINING FOR OLYMPIC ATHLETES

INTRODUCTION

The purpose of this Visualization Training Program is to define the concept of imagery as it applies to sport, to explain basic imagery technique and describe methods for improving imagery skill and specific ways to apply these imagery skills to your individual training.

"Imagery" is a very relevant to the athletes training, preparation and competition. Many olympic athletes and coaches are familar with and use imagery techniques for performance enhancement. The following information can provide practical methods for improving your athletic performance.

Author
Richard M. Fenker

Compiled By
Shirley Kay Durtschi
Research Assistant-Sport Psychology
U.S. Olympic Training Center

What Is Imagery?

Imagery is one language of the body/mind. Just as you often use words or self-talk as the basis for "thinking", you also frequently use images. Words or verbal languages are our most common means for communication with the external world (and for thinking about this world). Images are most often used for internal communication. Imagery can be visual (mental pictures or visualizations) but it can also be auditory (involving sounds), olfactory (involving smells or tastes), tactile (involving touch or feeling) or kinesthetic (involving your sense of body position or motion). An extremely vivid or clear image will usually involve more than one sensory modality. For example, if you were a swimmer, you might imagine a mental picture of yourself swimming in a competition and also hear the sounds of your teammates cheering, feel the coolness of the water against your skin, have an acute kinesthetic awareness of your body as you execute a quick, smooth turn and finally, smell the odor of chlorine from the pool.

Imagery Is The Language Of Experience

Images are our mental versions of experience. Using imagery we can create, in vivid detail, a replay of a past experience; or, we can create a new event or experience in our mind. In sports, we are free to capture the image of another athlete's moves or form and build on this pattern to develop our own style.
Imagery guides much of an athlete's experience because it is a more efficient, complete language than self-talk. Try to describe how to execute a golf-swing, in detail, using words. You could write a book! Now repeat this task using imagery. An appropriate image communicates the same information in a few seconds!

How Does Imagery Work In Sports?

There are two theories to explain why imagining yourself performing your sport might have an impact on your actual swimming performance. First, when you imagine yourself performing an action, you are transmitting electrical impulses to the muscles involved in executing that action. These neuromuscular impulses will, for vivid imagery, have the same form or pattern as the impulses generated when you are actually performing (but be somewhat weaker). Thus, mental rehearsal of an action may be strengthening the neural pathways involved when these movements are performed. Second, images may be an efficient way of coding or "representing" instructions for movement. In the example given above, forming an image of a correct golf-swing gives a simpler yet more complete description than is possible with words.

What If My Imagery Is Very Poor?

First, remember that poor visual imagery doesn't necessarily mean your other imagery senses are diminished. Some people rely more heavily on their auditory images or kinesthetic images than on visual imagery.

Imagery is a skill and like all skills can be improved with practice. Think of it as exercising a very special mental muscle. If your images are weak or unclear or lack quality in any modality, relax. Probably what you need most is practice, using some of the exercises given in this brochure. Your coach or sports psychologist can also help by providing training materials for enhancing imagery skills. Learning to talk took some time, so will learning to develop vivid imagery. But, you can do it with practice.

What's The Most Effective Way To Use Imagery?

No one knows for certain, but current research suggests the following guidelines:

1. Use a combination of imagery practice with actual practice for the best results.

2. The most effective types of imagery probably depend on skill level. Beginning and intermediate athletes may want to focus on basic skills while more advanced athletes may work on physiological controls, problem solving or competition scenarios.

3. Cassette tape programs can be useful for improving imagery skills. Commercially produced tapes or other tapes made for a group are probably less effective than tapes designed for individual athletes.

4. Choose a quiet place where you are not likely to be disturbed for practicing imagery. Using one of the trance techniques discussed below will also enhance the effectiveness of your imagery.

Imagery With Relaxation, Meditation Or Hypnosis?

Imagery practice will generally be more effective when done in conjunction with a "trance" technique such as self-hypnosis, relaxation, meditation or hypnosis.

There are three reasons for the increased effectiveness. First, "trance" techniques are methods for focusing attention or improving concentration. And using any of these methods insures that your perception of the images will be enhanced. Second, using one of these techniques is likely to improve the quality or vividness of your imagery. The "letting go" that takes place during relaxation or self-hypnosis is conducive to the production of imagery. Finally, trance states have a non-evaluative quality that keeps your everyday, rational mind from censoring the ideas or images that you create. For example, if you choose to imagine that your feet and legs are made of plastic and you can leap high over the other players on the basketball court, your logical mind might reject these images as irrational; but, in trance, anything you imagine is possible and can be believed. There are no limits to you imagination!

USING IMAGERY SKILLS IN SPORTS

This section describes the major uses of imagery in training and competition for athletes.

Mental Rehearsal

Imagery is used to practice some aspect of training or competition in the mind. The images might relate to: specific skills the athlete needs to master, making well-rehearsed skills more automatic or performing in competitions.

Skill Learning

Imagery is used to teach basic skills. Films of highly skilled athletes demonstrating the activity or examples given by the coach might be used as "models" for the athletes.

Competition Programs

Imagery is used to rehearse a detailed program for a future competition. A gymnast might imagine where the next meet will be held, the events in which he or she will be competing, and the details of the "plan" for managing each event. How will the gymnast get ready? What physical and mental details will be monitored? How will the competition be started, experienced and then finished? Will there be a strategy for dealing with key opponents, for being behind or ahead in points or for handling pain or distractions?

Problem Solving

It is very difficult to solve a problem unless you imagine that you can solve it. Imagery can be used to rehearse crucial problem situations and to find solutions.

Unexpected Contingencies

Athletes are often confronted with new and different situations that require adjustments. Imagining these situations in advance and making the required mental adjustments increases the chance that the appropriate response will be made when it counts.

Controlling Emotional Or Peak States

Each athlete has an image of himself or herself performing at a peak level. The thoughts, emotions and body feelings associated with this image are part of the athlete's "peak state". By knowing which physical or mental factors are associated with "peak" and then creating images that describe (accurately or symbolically) these factors, an athlete can gain some degree of control over the performance states. For example, one championship sprinter uses the images "relaxed, focused as a beam of light and gliding smooth and strong down the track" to describe her ideal performance state.

Controlling Anxiety, Tension Or Stress

Imagery can be used to prepare for situations where an athlete is likely to feel tension or stress. During rehearsal, a tension-producing situation is imagined and then "defused" by creating the appropriate corrective image. For example, excessive tension associated with the start of a race might be linked with the "image" of a simple breathing exercise (the exercise would be practiced as part of the imagery experience) that produces sensations of relaxation or control.

Imagery For Centering, Focusing Attention, Blocking Distractions, Managing Pain Or Energizing

There are many other types of imagery programs that can be use to achieve these and other specific goals. When an athlete experiences difficulty with attention, distractions, pain or other problems it may be valuable to design a program specifically for that problem. The appropriate imagery in each case depends to a large degree on the individual. Usually, however, such a program would involve: creating imagery of the problem situation (eg., experiencing pain); creating an appropriate "image" to handle the problem (eg., imagining that with each step some of the pain is pushed away from the body and left behind on the track); and finally, rehearsing the use of this image in practice or competition.

Imagery For Goals

Athletes can "program" their goals by using imagery to "see", "hear" or "feel" themselves achieving the goal. The imagery can be very realistic (a diver imagines her routine off the high platform--getting set, executing the dive perfectly, feeling the coolness of the water and watching the judges award a high score) or very unrealistic and exaggerated (a discus thrower imagines himself as superman lifting the discus without effort and hurling it out of sight).

GOAL VERSUS PROCESS IMAGERY

Should athletes use "goal oriented imagery" or "process" imagery in a performance enhancement program? Both types of imagery can be very valuable. Goal-oriented images of winning races, receiving medals or breaking records can be very effective images when used before or during workouts or before competitions. Such images which remind an athlete of his or her goals can be potent incentives for keeping energetic, attentive and motivated during practice. And, rehearsing one's goals before a competition can also have a motivational and energizing effect. But during the competition, any concern over goals or outcomes is likely to be distracting rather than helpful. Thus, for most athletes it is better to avoid goal-oriented imagery during competition. Images of winning or setting records can shift quickly into concerns over losing or slow times and in either case such images are likely to be distracting and to interfere with performance.

Process imagery is focused on the performance itself. What sights, feelings and sounds are appropriate for your best performances? What images are necessary for you to ski your fastest, making the quickest turns and eliminating all distracting thoughts? Each athlete has his or her own appropriate "process" images that guide performance by keeping the mind focused and "into" the competition. Process imagery can be used in practice but it is especially important during competition because it keeps the athlete's mind on the key elements of performance instead of winning, losing, how well opponents are doing, the crowd and other extraneous factors.

In general, rely on goal-oriented imagery as a motivator in practice or before a meet and use process imagery while you are performing.

SHOULD IMAGERY BE REALISTIC OR EXAGGERATED?

Researchers have not answered this question in a definitive manner. Another version of this question is to ask, "Should Goals be Realistic or Unrealistic"? The answer in both cases depends to a great deal on the psychological character of the athlete. Unrealistic goals or images are likely to be a problem for athletes who become frustrated or discouraged when these goals are not achieved. For other athletes, setting extremely high goals, even when unattainable, acts as an incentive or motivator to perform their best.

Each athlete or coach needs to make this judgment individually. If unrealistic goals or goal-oriented images lead to frustration, then set more realistic goals. If such images are effective as motivators, then use them.

Exaggerated images are often used by athletes to describe "how" they want to perform, not what goal they are seeking. Examples might be images of a boxer pummeling his opponent like a punching bag, of a high jumper who can fly over the bar and soar to an indefinite height, or of a tennis player who has eyes on her racquet and can "see" from the racquet's perspective where to send the ball. Such images communicate in a symbolic manner, a clear message about what is expected. Many athletes have used unrealistic goals and exaggerated imagery successfully. Select images that work for you.

GUIDE FOR INCLUDING IMAGERY IN TRAINING PROGRAMS

1. Imagery is like any other skill: it must be practiced regularly to be effective.

2. Imagery training should be included as part of the athlete's regular practice routine. If it is added as an "extra" to be done outside practice, this sends a clear message about its priority relative to physical practice.

3. Imagery can be used effectively before, during and after physical practice.

   (a) "Before" sessions can be used to prepare athletes for what will happen in practice or to create a particular mental state.

   (b) "During" sessions can reinforce how imagery might be used during an actual competition or help athletes learn new skills or strategies.

   (c) "After" sessions can help emphasize points made during the day or give athletes the chance to recover after a strenuous practice.

4. General imagery skills such as rehearsal or concentration can often be taught in a group setting. Specific skills such as problem solving or developing competition scenarios are best done individually.

5. The evaluation of athletes' imagery skills is a good starting point for an imagery program. The feedback will help build interest in the program and motivate individuals to participate fully.

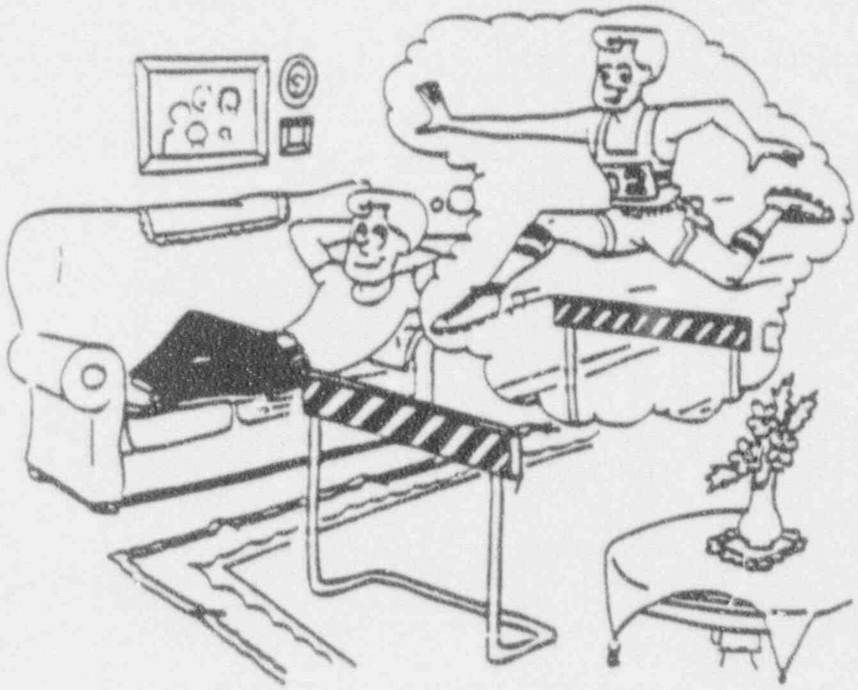
6. Regard imagery as simply another useful skill in an athlete's repertoire of physical and mental skills. For some people it will be of enormous benefit, while for others it will have little effect. Used properly, imagery training is a powerful tool for integrating the body and mind in sports.

ADDITIONAL INFORMATION SOURCES

Suinn, Richard, Seven Steps To Peak Performance.

Syer, John and Christopher Connolly, Sporting Body, Sporting Mind, Cambridge University Press, New York, 1984.

Williams, Jean (Ed.), Applied Sports Psychology, Mayfield Publishing, Palo Alto, 1986.



MAKING YOUR OWN VISUALIZATION TAPE

This Handout includes:

I. Purpose of Visualization Tapes.
II. Preparing to make a Tape.
III. How to structure your own Tape.

I. Purpose of Visualization Tapes

Virtually all successful athletes use mental preparation before competitions. For some athletes, the preparation is as simple as thinking about general strategies for the upcoming competition. Many elite athletes, however, take a more detailed and organized approach to mental preparation, and visualization is often a central part of that pre-competition training. Unlike fantasizing or wishful thinking, visualization is a purposeful and organized use of your mind to improve performance. It requires time and practice, and it is not a substitute for physical training.

If you have had problems with confidence, or have felt that your mind let your body down in a tough competition, then visualization could be very useful for you. Visualization helps you to prepare for competition situations, see yourself being successful, and maintain confidence in yourself. By rehearsing a successful performance, you can anticipate obstacles, recognize the opportunity to use your strengths, and stay calmer during a performance situation.

II. Preparations For Making A Visualization Tape

What You Need:

SCRIPT- Your tape will work best if you have thought carefully about what you want to include and how you want to present it. Like a carefully planned competition strategy, your tape should be detailed and organized. When visualizing, it is important that you make the experience as real as possible, and that you involve all of your senses, just like you would when competing. It is much easier to make a thorough tape if you write down a "script" ahead of time. Before making your script, you need to think about a few aspects of the upcoming competition:

1) Your Strengths- One of the most essential parts of a visualization tape is a focus upon your strengths as an athlete. In preparation, write down a list of your strengths as a competitive athlete:

* What parts of your competition performance have helped you in the past? What abilities do you have that you feel confident about?
* Are you stronger than most opponents?
* Have you trained hard so that you have a good fitness level?
* Are you usually quicker or faster than opponents?
* Are you mentally tough and good at enduring pain?
* Are you steady?
* Are you aggressive? Have you developed specific skills?

respect. If you do use someone else's voice, it is especially important to make the script fairly detailed and personal.

<u>What to say on the tape</u>? The tape should be tailored to your own sport and sometimes to a specific competition, but we suggest the following four sections as general guidelines:

1) The relaxation phase- your tape should begin with a brief relaxation portion which allows you to shut out outside thoughts and nervousness. One good and simple exercise is to focus on taking long slow deep breaths from your diaphragm, holding the breath a bit, then slowly exhaling. Do this a number of times and you should notice feeling more relaxed almost immediately. If you use another relaxation technique, write that into the script here.

2) Pre-Event- Going back to images you collected, try to simulate a very positive pre-event state of mind and body. As you notice your competitors, you feel calm and confident. You feel good about your preparation, and as you take in your environment, you know you will have a good performance. You remind yourself of the two or three things you need to focus on in this competition, and begin to think about your strengths and why you will do well

3) Event- As you visualize the event beginning, you recognize that you feel calm and powerful and in control of your body and mind. You focus on the two or three things that are keys to your success, and you see yourself successfully doing them. ((If you get too nervous or start to see negative or unsuccessful images, just stop for a second and go back to some relaxation exercises. Don't get mad at yourself, because it is natural to have to work through the rough spots. Remember, if you get nervous or get negative images during visualization, it is likely that you would during actual competition. By working through this now, you will have an advantage over others who must deal with this during competition)) When you can see yourself performing successfully, follow your performance and notice how good it feels to be doing well. Try to use all your senses and feel the different sensations as you begin to recognize that you are succeeding. You stay calm and relaxed and keep up your strategy until the very end. You are not cocky, but you are not defensive: you know that maintaining your strategy will enable you to succeed.

4) Post-Event- After seeing yourself perform successfully, take a moment to feel what success is like. Take in your environment, and notice everything. Enjoy the success for a while. After you have a good clear image of what it feels like to succeed, go back

to a few deep breathing/relaxation exercises, and gradually exit from the visualization, gradually becoming more alert. Try to determine how you feel. You should be calm and relaxed. If you are very aroused and excited, you should do a few more minutes of relaxation. It is sometimes helpful to replay your visualization, because you may notice aspects of your performance that you don't notice in the heat of actual competition. If there is something you don't like about your performance in the visualization, write a corrective statement into the script and in your next visualization you can modify the behavior.


START
START
STOP
START

SUGGESTIONS FOR MAKING A PRECOMPETITION TAPE

Many athletes like to get in the right frame of mind for competition by listening to a tape before they compete. Tapes can be played a week before competition, the night before competition, and now, with the invention of Walkmans, they can be listened to AT competition. Some athletes use tapes that have their favorite psyching-up or relaxing music on them, while others listen to tapes that contain suggestions about important things to concentrate on before competition. Some tapes contain both music AND instructions.

You can make a tape, or tapes, for yourself quite easily. First, decide if you want a music tape or a suggestions tape, or both. If you want a music tape, decide what music makes you FEEL BEST before competition. Remember, you may be listening to this tape on the team bus, or perhaps at the track or the swimming pool, wherever your sport event takes place.

Next get a copy of the music you want to listen to, whether it is on a record, CD, or cassette. An athlete about to compete in a boxing event might make a tape with the themes from "Rocky", "Top Gun", and "Star Wars". This tape will get him psyched up. Another athlete who is tense before a shooting competition might make a tape with music from "Chariots of Fire", "Pachabel's Canon", and some favorite classical music. This tape will help calm the athlete down.

A suggestions tape will help you get in the right state of mind for your best performance. It will help you cue in on the things you need to be thinking about before competition. If you make your own tape, you can personalize it. This will generally be more effective than listening to a mass-market tape that is not attuned to your special and unique circumstances.

First, decide whose voice you want to be listening to. Some athletes prefer listening to themselves, but others much prefer someone else to be speaking. If you select someone else, it should be someone who gives you confidence when you listen to them - a coach, a close companion, or your sport psychologist. Next, remember to try your tapes out before you go to competition. Try them out at practices, and under simulated competition situations, so that you can be sure they have the right feel, and that their timing and pacing is right. If you need to modify the tape, you will have plenty of time that way.

SAMPLE TAPE

1. Relaxing music.

2. Relaxation instructions. These should be fairly brief unless this is just a relaxation tape. An example of a complete relaxation script is provided in another handout. Concentrate on deep breathing and relaxing the muscles.

3. Self-suggestions. Some examples include:
   * You have every reason to believe in yourself, as a competitor and as a person.
   * You are well prepared. You are physically ready. You are mentally ready.
   * Remember your goals. You have the ability and the determination to reach them.
   * You are fully prepared to dig deep within yourself to meet these challenges.
   * Focus on what YOU have to do and on what YOU are going to do. Nothing and nobody else matters.
   * Focus on WHAT IS WITHIN YOUR CONTROL.

4. Final suggestions. Go over the key points that you need to focus on. These cues may be thoughts that have helped you to attain your previous performance bests. For example:
   * Be powerful at the start.
   * Nothing will stop you today.
   * Concentrate on good extension.
   * Just focus on your body and let your performance come naturally.

ADDITIONAL SUGGESTIONS

These self-suggestions have been found helpful by other athletes. Some of them may work for you:

* The challenge before me will require an all-out effort. I am fully prepared for the challenge - I will push my limits.
* Because of the physical preparation I have done and the mental plan I have developed, I have an advantage.
* I can draw upon all my resources.
* I will be as I choose to be.
* I choose to be my best.
* I choose to be the best.
* As I prepare to start, I will become completely concentrated and absorbed by the task. I will feel energized and in control. I will follow my plan. Nothing will disturb me or distract me.
* During training for this event, I prepared my body and mind for a superior performance.
* I planned and prepared my precompetition warm-up to create an ideal feeling state for my performance.
* I developed a refined competition plan to take me to my goal.
* I simulated my performance physically, mentally, and tactically. So I am ready.
* I have every reason to feel confident. I feel good. I am better than anyone here. I am the best. I am fully prepared. I am confident in my preparation, and I am completely determined to achieve my goal. People are expecting me to do well, and I am going to show them what I can do.

These are some suggestions if someone else is reading the tape for you:

* Stand tall! Let the champion within you push forward.
* You are already starting to feel the way you want to feel, and you will perform the way you want to perform.
* You are and will remain mentally calm, alert, and absolutely determined throughout.
* You have a well-practiced plan that will unfold as you have imagined, hoped, and planned.
* If you experience any anxiety before the start, it will be transformed into the explosive power of your start.
* You are capable of adapting and refocusing in the face of any obstacle.
* You can control what you do.
* In the competition you will follow your plan, and you will be focused and in control throughout.
* You can control your own destiny.

LOST PRE-EVENT TAPES

What happens if you prepare a pre-event tape and then can't find it when you need it at the competition site? Treat it as you would a piece of your equipment, and you will not likely lose it. In case of loss or breakdown, perhaps it would be worthwhile to carry a copy of your tape or a written copy of your pre-event self-suggestions as a backup. If you end up not having a copy of your tape to listen to or a copy of your self-suggestion to read to yourself, sit down quietly in your own space and think about how you want to feel. Try to recall your major self-suggestions. That should not pose a problem, because as a result of your previous experience with the tape, your self-suggestions should be well learned by this time.

RE-EXAMINATION OF THE ROLE OF ASSESSMENT IN PROMOTING STUDENT LEARNING

By

James E. Bruno

There are three major types of policy analysis studies that are associated with issues that involve student testing and assessment. These are the "**why**" of assessment, the "**how**" of assessment and the "**what**" of assessment.

Why societies assess their young is embedded in the collective culture of mankind or as Carl Jung would say the "collective unconscious" of man. The rite of passage of youth, the Bar Mitzvah, the confirmation, the SAT, etc., always seems to entail some sort of "ordeal" or rite or performance based test for the youth of a culture. Even the ancient Greeks were keenly aware of the need for assessment. When the Sphinx asked Oedipus "what walks with four legs then two legs then three legs" and Oedipus answered man, he was "judged" by the Sphinx as being "king" material. History's first Criterion Referenced Test was thus recorded. Over the millennia of man, only the format of testing -- oral, performance, written-essay, and multiple choice has changed. The purposes of testing and assessment seem to remain the same regardless of culture -- sorting of individuals.

A. Why We Test

In a modern twentieth century, industrial state, the "**why**" of testing in the classroom has two important purposes. These are:

(1) **Summative Evaluation** -- judgment, selection or sorting of students

(2) **Formative Evaluation** -- assessment of a student to promote classroom learning.

Sometimes tests are used in the classrom for both summative and formative evaluation purposes. Naturally it is the latter purpose of testing and assessment that is perhaps most important for promoting student learning. Assessments for formative evaluation have to be concerned about the natural tendency of students to **forget** material (previously taught) and the propensity of students to **learn new material** (to be taught). Both **forgetting** old information and **learning** new information are strongly related to the expressed confidence that students have in their current information base -- i.e. their levels of reliable (accurate and confident) information. See Figure 1 for an analysis of how learning new information (i.e. percents) is based on the confidence in previously learned information (i.e. decimals). Also note how previously learned information is forgotten over time in direct relationship to expressed confidence in that information.

Thus "why" we assess students, at least from the **classroom teacher perspective**, is basically twofold.

(1) Judgement, Selection and Sorting (grades and placement) -- **Summative Evaluation**

(2) Enhancing Learning of new information and minimizing the forgetting of old information -- **Formative Evaluation**

From the **student (and parent) perspective** the "why" of testing and assessment includes:

(1) Insuring that the student is "on line" in terms of meeting the educational objectives of the family and culture

(2) Articulating the school's instructional program to students (and parents) so that students can help themselves learn.

From the **school administrative perspective** the "why" of testing includes:

(1) Resource allocation (micro level) to promote the effectiveness of the instructional program

(2) Teacher accountability and assignment

(3) Curriculum alignment and planning.

From the **state perspective**, the "why" of testing includes:

(1) Educational accountability

(2) Resource allocation (macro level) to promote the state's instructional objectives.

B. How We Test

The "**how**" we assess students is based on two major perspectives to assessment. Will the evaluation be **objectively** scored (multiple choice, low cost, machine scoring) or **subjectively** scored (performance based, portfolio analysis, high cost teacher evaluation)? There are important reasons why **objective** testing is presently favored over **subjective** testing. Some of these reasons include economic (cost and viability), Legal (bias and due process), and Educational (domain of skills, information for instructional decision making) considerations. In general objective testing is preferred over subjective assessment because it can assess a wider domain of skills, eliminates evaluator bias from the assessment process and most important is economically viable for the educational agency to perform on a mass scale.

Presently there are **three** types of objectively scored tests that are available to the classroom teacher. These are the **Norm** Referenced Test (NRT), the **Criterion** Referenced Test (CRT), the **Information** Referenced Test (IRT). The following table depicts some of the main characteristics of each type of test.

Type	Mechanics	Type of Objective Scoring
Norm Referenced Test	Student Score is Referenced to a Normal Distribution Curve	Right or Wrong (one dimensional and non-reproducible)
Criterion Reference Test	Student Score is Referenced to a Selection Standard Score	Right or Wrong (one dimensional and non-reproducible)
Information Referenced Test	Student Score is Referenced to Information Standard of attainment	MCW-APM (two dimensional and reproducible)

The **Norm** Referenced Test such as the Stanford Achievement Test (SAT), and the Comprehensive Test of Basic Skills (CTBS), is typically used for summative evaluation. They are notoriously sensitive to student guessing, test preparation and culture bias. These limitations make them nearly useless sources of information for the teacher for promoting classroom learning. The fact that the scoring system used on a Norm referenced test is **not reproducible**, since guessing is implicitly encouraged, further limits its practical use in classroom formative evaluation. Finally the tautology of assuming that the **normal** distribution model of attainment is perfect, then designing test items to fit that model, underscores the **pseudoscientific** nature of the Norm Referenced Test for use in Selection or Summative evaluation. In short, the Norm Referenced Test is generally considered by teachers to be useless for measuring anything but the narrowest, instructionally irrelevant set of skills.

The **Criterion** Referenced Test (CRT) such as mathematics or language placement test are far more "in line" with teacher needs for instructional decision making and used in promoting

classroom learning. Unfortunately, due to the one dimensional system of evaluating (Right or Wrong) or scoring these types of tests, they are also extremely sensitive to guessing. In addition there is a strong inverse relationship between the selection standard or criterion score used on the CRT and the number of false positives generated by the assessment process.

The **Information** Referenced Test is a hybrid NRT and CRT. Like a CRT, the main emphasis is for classroom formative evaluation and like an NRT the score is referenced to some fixed standard or model of attainment The Information Referenced model of attainment evaluates students on a standard of being informed, part informed, uninformed and misinformed. The scoring system used in an IRT is also a **reproducible** system. With a **reproducible** test scoring system the maximum score for the student is obtained **if and only if** the student does not guess, but answers each test item on the assessment with his or her actual information. The IRT procedure is fully **optically scannable** (see Figure 2) and generates an **individual education plan** (IEP) for formative evaluation of each student assessed (see Figure 3). A **class information needs profile** (CINP) **(see** Figure 4) and a **school information needs profile** (SINP) (see Figure 5) are also produced with the IRT procedure.

C. What Are the Limitations of Present Testing Practices

Policy studies that deal with the "**what**" of testing and assessment practices, tend to underscore the severe limitations of presently used CRT and NRT formats. The search for alternative testing and assessment formats such as the Information Referenced Testing (IRT) format and the Answer Until Correct format, are research efforts to address some of these concerns.

Here are just a few of some of the major concerns, as found in the research literature, with presently used CRT's and NRT's for classroom use.

- Teachers are not provided with information needed for instructional decision making.
- In many inner city schools, nearly 1/3 of the students test at or **below the random chance** or expectation level.
- The correction for guessing formula assumes that all wrong answers on a test were the result of random guesses. This formula thus over corrects and yields a **biased estimates downward** of actual student attainment.
- When corrections for guessing are not used the estimated of student attainment are a **biased estimate upwards** of actual student attainment. Students actually know less.

- The one dimensional Right or Wrong (RW) test scoring system is in reality only a recognition (and non recognition) system and promotes a student test score maximizing strategy that necessitates guessing. It is a **non-reproducible** test scoring system.
- There is **no partial credit** given with presently used multiple choice testing formats. This limitation leads to large disparities between the student test score maximizing and teacher conforming strategies. This disparity in strategies is inverse to student attainment levels.
- Because of the large amounts of misinformation, lack of information and partial information, presently used NRT's and CRT's scored with one dimensional R-W test scoring systems are **insensitive** or yield erroneous results when used with low and medium attaining students.
- Misinformation knowing something wrongly 1/2 + 1/3 = 2/5 cannot be assessed with present practices. Thus the important **pedagogical** strategy of reeducation for misinformation cannot be addressed.
- **The instructional leadership TRIAD** - student (and parent), teacher and school administrator are not part of the evaluation loop of instructional support.

- Current CRT and NRT testing practices assume that the **signal of instruction** is perfect and that all imperfections in instruction are assumed to be with the student. In essence, instructional validity is not considered nor is it assessed.
- The Normal Curve Assumption of ability, which is the foundation of the NRT format, is predicated on two erroneous propositions. These are that all student responses are certain (flipping a coin -- heads or tails -- right or wrong) and that the narrow domain covered by the test is an unambiguous measure of general student ability. Basically the normal curve model is assumed to be perfect and all imperfections to produce the normal curve model reside with the test item rather than the student or the model itself.
- There is by design a **ceiling and cellar effect** with grade level NRT. This limitation can and does lead to enormous distortions at each end of the distribution for assessing student attainment.
- There is no cumulative feedback curricula information provided to the teacher or school principal regarding common areas of misinformation and lack of information.

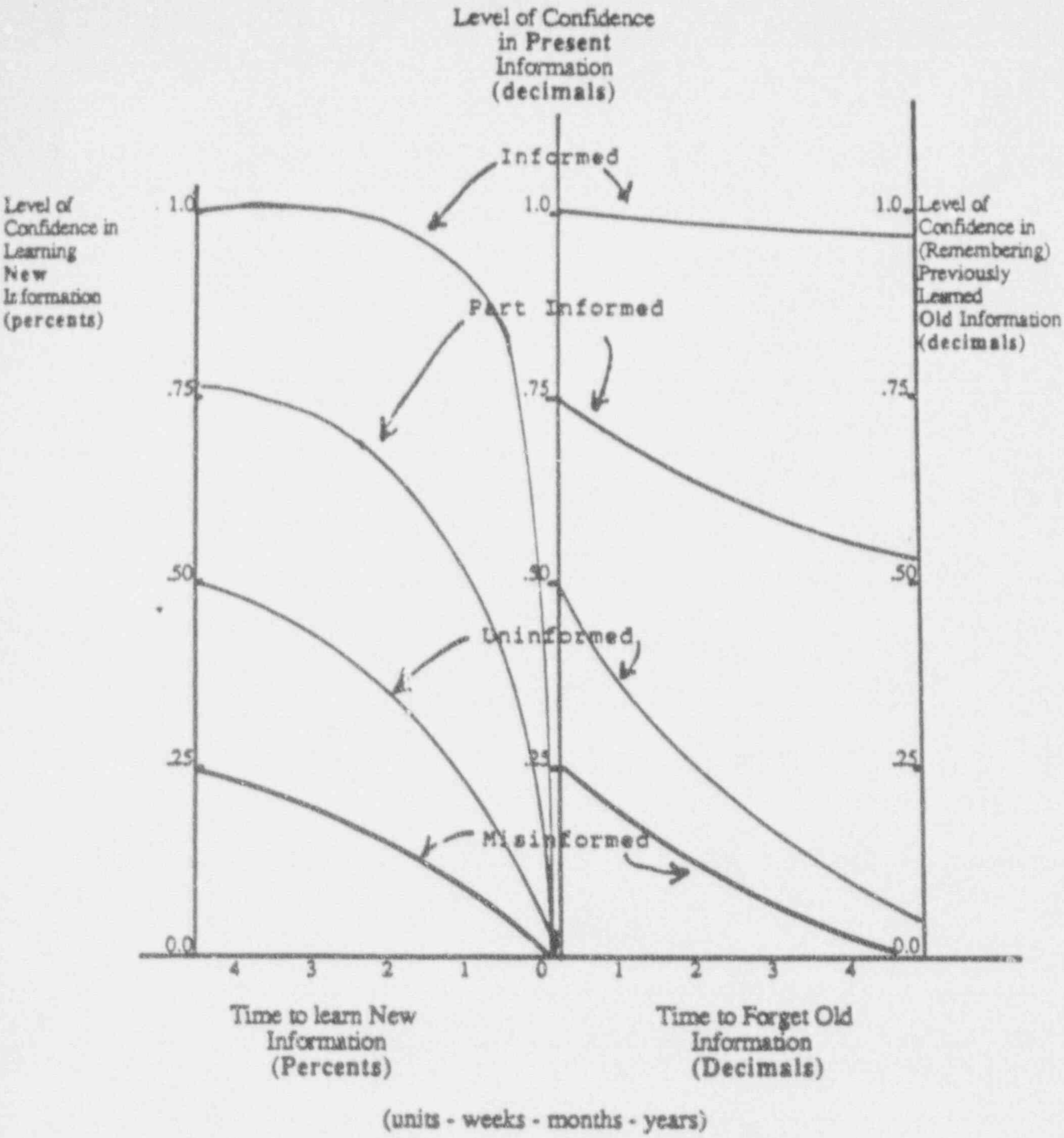
D. Conclusion

In conclusion there will always be problems with assessing student attainment in the classroom. Biochemistry has not advanced to the state where students will produce an enzyme when he or she has learned algebra. If this were ever the case teachers, by means of a simple blood test, could then clearly and unambiguously determine if actual learning had taken place. Until these advances in biochemistry come to fruition, all educational assessments will have to be "passive" or one step removed assessments.

What seems to be needed in education is a total re-examination of the basic philosophy of why we test. The philosophy presented here is that all testing should be used to promote learning. The way medical doctors use testing to promote the health of a patient, teachers should use testing to promote the learning of the child. The way medical testing is used to monitor health and suggest strategies for improving health, educational testing should monitor learning and suggest appropriate instructional interventions to enhance learning.

The design of the **Information Referenced Testing** concept is directed specifically at addressing the formative evaluation needs of an **individual** student and thus promoting learning. As other multiple choice objective formats and subjective testing formats become available, they should all be evaluated against the economic,legal as well as the educational standards for formative as well as summative evaluation. Unless an alternative to present multiple choice, CRT and NRT testing

practices occur, school reform will have no chance of succeeding. A "fix" on exactly what has to be reformed to improve learning will be missing and the important relationship between testing and learning left unaddressed. Information Referenced Testing is a significant step in filling the important **formative** evaluation niche in schooling and might provide the type of "sensitive" instrumentation that is needed to service the demands of the school reform movement.



Linkage Between Forgetting (old information) Learning (new information) and Confidence in present Information

Figure 1

Learning New Information and Forgetting Previously Learned Information as a Function of Expressed Confidence in Information

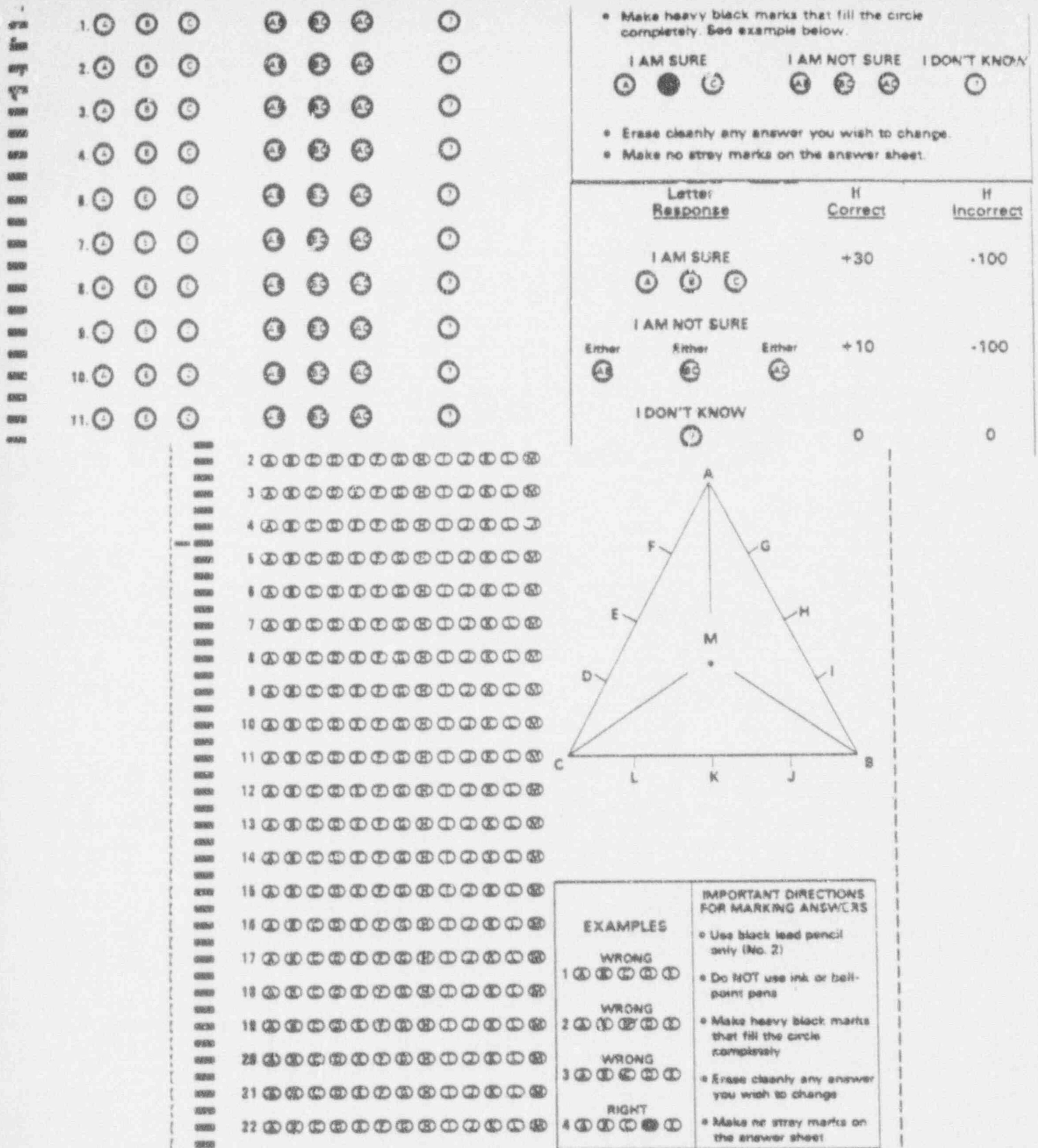


Figure 2

Optical Scan Sheet to Support

Two Dimensional Information Referenced Testing (IRT)

in the Classroom

IRT-produced individual education plan (IEP) to support information-referenced grouping of students

Formative Evaluation

Examinee Misinformation of Examination

Concepts where you were sure of an answer but were wrong

Have instructor explain why the answer was wrong that you thought was correct and why another answer was correct.

Test Item	(INF STATE)	Description Instructional Cross Reference
1	M	Addition—Decimals 1, 2, and 3 Places General Math, p. 94
4	M	Problem Solving (WP)—Multiplication 2 Step General Math, p. 32
3	M	Addition and Subtraction of Integers General Math, p. 62
7	M	Integers Number Lines General Math, p. 395
.	.	.
.	.	.
.	.	.

Examinee Uninformed (lacks information) Response

Concepts that you said you didn't know—have your instructor explain these concepts to you.

Test Item	(INF STATE)	Description Instructional Cross REference
10	U	Least Common Multiple General Math, p. 190
11	U	Problem Solving (WP)—Addition and Multiplication General Math, p. 208
13	U	Addition and Subtraction Mixed Numbers General Math, p. 190
15	U	Prime Factors Intro to Algebra, p. 58
.	.	.
.	.	.
.	.	.

Examinee Partially Informed Items on Examination

Concepts where you weren't sure of the answer—have your instructor review these concepts with you.

Test Item	(INF STATE)	Description Instructional Cross Reference
6	P	Problem Solving (WP)—Subtraction 1 Step General Math, p. 61
8	P	Calculating Percent of a Number General Math, p. 300
22	P	Division Decimals 2 and 3 Places General Math, p. 112
21	P	Calculating—Numerical Expressions Intro to Algebra, p. 2, p.44
.	.	.
.	.	.
.	.	.

Figure 3

Student Individual Education Plan (IEP)

Generated with the IRT Procedure

... IRT-produced class information profile to support educational grouping:
Instructional objectives for the teacher

Misinformed Examination Items for Your Class

Your students have wrong information in these concept areas—misinformed. Develop instructional materials to counteract these misconceptions. Follow this instruction with basic instruction in these areas.

Test Item	Percent of Class	Concept Description/Cross Reference
16	0.66	Numerical Expressions—Integers Intro. to Algebra, p.44
7	0.62	Integers—Number Lines General Math, p. 395
20	0.57	Least Common Multiple General Math, p. 130
34	0.56	Comparison—Rational Numbers General Math, p. 413
.	.	.
.	.	.
.	.	.

Uninformed Test Items for Your Class

Examinees generally lack information in these concept areas. Basic instruction is needed. Prepare basic instructional materials to teach these basic concepts.

Test Item	Percent of Class	Concept Description/Cross Reference Calculations
8	0.27	Calculations General Math, p. 300
1	0.23	Addition—Decimals 1, 2, & 3 Places General Math, p. 94
4	0.23	Problem Solving (WP)— Multiplication 2-step
.	.	.
.	.	.
.	.	.

Partly Informed Concept Areas for Your Class

Your examinees have incomplete or unstable information in these concept areas. A thorough review is needed in these areas.

Test Item	Percent of Class	Concept Description/Cross Reference
3	0.23	Subtraction (1 & 3 Decimal Places) General Math, p. 94
9	0.78	Multiplication—Fractions (Proper) General Math, p. 204
30	0.77	Conversion—Percent to a Decimal General Math, p. 376
.	.	.
.	.	.
.	.	.

Figure 4

Class Information Needs Profile (CINP)

Generated with the IRT Procedure

MISINFORMED EXAMINATION ITEMS FOR YOUR EXAMINEES
YOUR EXAMINEES HAVE WRONG INFORMATION IN THESE
CONCEPT AREAS—MISINFORMED

HAVE WORKSHOP COORDINATOR DEVELOP INSTRUCTIONAL MATERIALS
DEMONSTRATE MISCONCEPTIONS—FOLLOW THIS WITH ACCURATE
INFORMATION

ITEM	NUMBER	PERCENT	CONCEPT DESCRIPTION/CROSS REFERENCE
17	240.[illegible]	0.73	WORD PROBLEM—PERCENT GENERAL MATH P 314
9	[illegible]06.00	0.50	WHOLE NUMBER—FRACTIONS—LCD GENERAL MATH P 190
27	151.0[illegible]	0.5[illegible]	INTEGERS—ADD AND SUBTRACT GENERAL MATH P 394
34	177.00	0.53	FRACTION—COMPARING GENERAL MATH P 178
35	[illegible]09.00	0.47	PRE-ALGEBRA TRIANGLE FORMULAS INT ALGEBRA P 104
22	[illegible]50.00	0.45	PRE-ALGEBRA DISTRIBUTIVE PROPERTIES INTERMEDIATE ALGEBRA P 31
[illegible]	14[illegible].00	0.44	INTEGERS—ADD AND SUBTRACT GENERAL MATH P 394
23	[illegible]20.00	0.38	FRACTION—GRAPHS—CIRCLE GENERAL MATH P 369
32	[illegible]22.00	0.30	PRE-ALGEBRA USING GEOMETRIC FORMULAS INTERMEDIATE ALGEBRA P 37
6	[illegible]19.00	0.36	WORD PROBLEM—WHOLE NUMBER ESTIMATION GENERAL MATH P 269
11	116.00	0.34	FRACTION—ADD—SUBTRACTION MIXED GENERAL MATH P 179
28	113.00	0.34	PRE ALGEBRA—SEQUENCES AND PATTERNS INT ALGEBRA P 444
29	111.00	0.33	PRE ALGEBRA WORDS TO EQUATIONS INT ALGEBRA P 130
4	[illegible]08.	0.31	WHOLE NUMBER—DECIMAL ROUNDING GENERAL MATH P 17,120
24	104.0[illegible]	0.31	PRE ALGEBRA RADICALS,POWERS INT ALGEBRA P 453
5	93.00	0.28	WHOLE NUMBER—PRIME FACTORS INT ALGEBRA P 63
11	76.00	0.23	WORD PROBLEM—FRACTION MIXED GENERAL MATH P 190
7	76.00	0.23	DECIMAL,POWERS AND SQUARES GENERAL MATH P 266
16	76.00	0.23	PERCENT— % OF A NUMBER GENERAL MATH P 300
37	71.00	0.21	PRE-ALGEBRA—EVALUATE ALGEBRAIC EXPRESSIONS INT ALGEBRA P 5
1	70.00	0.21	DECIMAL ADDITION—HORIZONTAL GENERAL MATH P 84

UNINFORMED TEST ITEMS ON THE EXAMINATION
EXAMINEES GENERALLY LACK INFORMATION IN THESE
CONCEPT AREAS

BASIC INSTRUCTIONAL MATERIALS ARE NEEDED IN THIS AREA

ITEM	NUMBER	PERCENT	CONCEPT DESCRIPTION/CROSS REFERENCE
36	186.00	0.30	PRE-ALGEBRA SIMPLIFY ALGEBRAIC FRACTIONS

Figure 5

School Information Needs Profile (SINP)

Generated with the IRT Procedure