A Brief Introduction to The Word Associate Test

William S. Verplanck,
University of Tennessee, Knoxville

An examination format assessing the intraverbal repertoire of individuals in psychology is described and results using it reported. The Associate Test is easy to prepare, to take, and to grade. Its reliability measures are satisfactory; its ability to predict later behavior is reported upon. The Associate Test is computer friendly, and its methods can be applied for examination in any field, and at any level.

The Word Associate Test (WAT) (or Short-Short Answer Test) is a format for preparing brief tests designed to demonstrate the student's command of the critical terms of specific areas of study. It provides a measure of the student's vocabulary in specialized and technical areas. That is, it samples the minimal intraverbal repertoires of the individual.

The Word Associate Test has been developed over some 30 years of research.¹ This work was instigated by dissatisfaction with the multiple-choice test, whose scores are patently associated with specific content-independent verbal skills and with well-shaped moments for guessing.

The primary objective of the WAT is the rapid examination and grading of large numbers of students in one or another of a broad variety of areas (Jennings, 1968; Saeger, 1975), and at a number of levels of achievement. It was not developed to find

how people think, how original² they are, how well they can explain something, or write, much less to predict academic success, life achievement, or aptitude for something or other. The command of specified vocabularies is, however, necessary for all these performances.

The WAT asks only one question, which has many paraphrases in ordinary language: "How successfully has he/she studied?", "What words can the teacher or writer use and expect the students to be able to deal with at an acceptable level?", and loosely, "How much does the student know?" It does so without the use of the prompts that appear in multiple-choice, true-false, matching, and fill-in-the-blanks examination items. Incorrect prompts may indeed be added to the examinee's intraverbal repertoire.

The WAT may be termed "robust": instructions to guess do not increase scores; instructions "Do NOT guess," however, decrease them.

The WAT demonstrates the examinee's minimum level of command of a selected vocabulary. The WAT does not enable students to demonstrate their full knowledge.

The writer would like to acknowledge gratefully the work of J.R. Caplan, R. Fleischer, A. Gordon, W.N. Jennings, R.N. Krueger, R.A. Ramos, W. Saeger, and J. Shelnutt. P.J. Lazzara and R.D. Kinser assisted in summarizing this research and in assembling the appendices.

¹The progressive development of the Associate Test and its methodology into its present form have been reported at meetings of The Psychonomic Society (6), of the Southeast Psychol. Ass'n. (6), and of the American Education Research Ass'n. More detailed reports have appeared in 1 Honors Paper, 2 MA Theses, and 2 Doctoral Dissertations. A listing of these papers may be requested of the author.

²Jennings (1966) has nevertheless shown that there is a relationship between word-associate measures and evaluation of the "originality" shown in compositions written for the instructors in The English Department at UTK. These instructors had no information whatever on the WAT, much less of each student's performance.

For this, a further WAT, presenting a more advanced technical vocabulary should be taken. In essence, a WAT can be prepared for any level of study, from kindergarten through the doctorate, in any field (with the possible exception of some branches of mathematics).

Although the test was developed before computer technology had reached its present level of sophistication and accessibility, it was preadapted for computer administration, grading, and presentation of results. Computer technology will enable a student to take an Associate Test on any subject at any level of achievement at any time at any terminal with immediate scoring against national norms.

It may now be applied to evaluate the achievement of individual students or specific classes and the effectiveness of various teaching methods, or of instructors. (Jennings, 1968; Saeger, 1975; Shelnutt, 1969). The description which follows gives the present status of the Word Associate Test. The appendices present sets of data relevant to the use of the test and to how it works in practice. An extended monograph on the WAT is now in preparation, and is to be published by the Cambridge Center for Behavioral Studies. This will report on the research on the WAT carried out to date, and will indicate its potential as a research instrument as well as a method for examination of substantive command of subject matter. Users will evaluate its face validity on the basis of their experience with its use and with the students who have taken it.

Given the current state of understanding of relevant criteria and of "psychometric validity," the psychometric validity of the WAT is no better or worse than that of any other form of examination, whether essay, take-home, true-false, or multiple choice. It can and will be compared with others at such time as acceptable sets of well-defined criteria can be agreed upon.

Although the WAT has been independently developed, the readers of Rosanoff (1918), Deese and Loftus will find it of both practical and theoretical importance. The reader will also recognize that "Word

Associate Test" is perhaps too restricted a name. "Associate Test" will do. Besides words, names, and dates, one may also use equations, curves, chemical formulas, diagrams and other visual material as stimulus items in preparing examinations using this design, and examinees may respond by drawing pictures and sketching graphs or diagrams, as some have done.

When the response given to the stimulus word or term is verbal, rather than a diagram, sketch, or graph, the associate may be considered a *primary* or *first-order intraverbal*, one that is given directly to the stimulus. These are the responses most likely to appear in sentences written when the examinee is asked to write brief essays on the term. In these answers, the first term to appear is most often the stimulus term itself — in this context, a notate, or tact.

The possibility of the use of stimuli other than written or spoken words, such as pictures, diagrams, sounds, and so on, is one reason why the term "associate," rather than "intraverbal behavior" (Skinner, 1957), is used. Certainly the specific methodology presented deals with intraverbal behavior. But the methodology lends itself equally well to the investigation of tacts, as distinguished from "intraverbal responses." Word associates are intraverbal behaviors only when the stimuli themselves are words or phrases; otherwise, they are tacts. In Skinner's *Verbal Behavior*, the terms "tact" and "intraverbal response" have extensive connotations, which may or may not be relevant to the verbal behaviors reported on here. A second reason for omitting these terms is that psychologists unfamiliar with or ignorant of the vocabulary of behavior analysis may not be able to take advantage of this methodology for examining what they would not refer to as either "verbal repertory" or "intraverbal behavior," even though this association procedure assesses the intraverbal repertoire. A vocabulary serves better as a bridge between divergent systematic viewpoints than as a barrier between them.

Finally, and by no means least impor-

tant, students seem to like it (Attkisson & Snyder, 1975).

THE WORD ASSOCIATE TEST: PREPARATION, ADMINISTRATION, GRADING

The Word Associate Test has been developed to fill the need for a test that can be easily and "objectively" graded, that can sample broadly the content of a course, that can be readily prepared and that avoids the multiple choice format. The format of the test is straightforward: stimulus terms—key words—are chosen from the material on which a class is to be tested (e.g., terms in the lecturer's notes, in the body, index or glossary of a text). Examinees are given the opportunity to demonstrate their familiarity and grasp of each term by responding to it appropriately. The stimulus terms selected are arrayed vertically; the number of responses required is chosen by the examiner. Instructions are brief; grading is similarly uncomplicated. Each response is graded on a scale chosen by the instructor. The distribution of summed scores is the basis on which letter grades can be assigned. This presentation will provide the readers with the information they will need if they wish to use this examination format. A monograph summarizing the findings of research on this method of examination is in preparation. Appendices present sample WATs, completed and graded, as well as other findings.

INTRODUCING THE WAT TO STUDENTS

Before giving a WAT ("The Short-short Answer Test") as a grade determining tool, the instructor should give the students one or more brief practice sessions. These serve as a quick review for the students, showing them what they need to study more, as well as giving them an opportunity to know how the method works on a test.

CONSTRUCTION

Instructions

The instructions below have proven easy to follow. A briefer form may be used at

the top of each page of an examination. (See Appendices.)

In each space provided for each item write a word, date, phrase, term, name, or concept relevantly associated to it in the subject matter of this course. Do not repeat. You may review your answers before passing your paper in — you'll probably be able to fill in some of the spaces you left blank.

References to grading need not occur in the instructions. Dry-runs, advisable before the first administration, serve to desensitize students to the novelty of the exam (new kinds of tests are always scary), as well as to eliminate the possibility that a student has misconstrued the instructions.

Choice of Stimulus Terms

The number of stimulus terms in a given test will be determined by how broad a sample the instructor wishes to examine on, how many responses he or she wants, and on the time allotted for the test. When two responses are required, students take approximately one minute per item to take the test.

The instructor should choose the terms that he or she thinks the students should know something about from lectures, discussions, or other assigned readings. Stimulus terms can and should sample a wide range of the materials covered in the class. Tables of contents, indexes of textbooks, and glossaries are especially convenient sources. Boneau's paper (1990) is a useful source.

Overly general or vague terms should be avoided; anybody can give acceptable associates to, say, "motivation," or "motherhood." Also avoid names easily identified generically, e.g., "Schultz," which will elicit "German psychologist." Related terms, i.e., words that would be acceptable as responses to one another, may both be included, but should be presented distantly from one another on the examination sheet.

The wide range of sources for stimulus items allows the test to be finely tuned to the content and level of the course. The test itself can be a very specific probe of student knowledge in given subareas. Eventually each instruc-

tor will be able to assemble a 'master list' of terms that can be given to students as study guides—the concepts of which they should be in command.

Appendix A lists, in alphabetic order, approximately 200 items from a set of over 2,000 psychological terms that have been used in first courses in General Psychology and Introduction to Biological Psychology at U.T. Knoxville.

Format

Terms are arranged vertically, isolated from one another by lines. This format discourages sentence writing, and increases the likelihood that the second, third, and fourth associates are given to the stimulus term, rather than to the immediately previous written response.

The student is usually required to produce four associates for each stimulus term, although this number can vary as circumstances require. Research indicates that four associates are ample to provide discriminatory power for evaluating student knowledge. Fact is, two associates prove to be enough to demonstrate basic knowledge. A test including 64 two-response items can usually be completed by students within an hour.

Appendix B presents a number of test formats, some completed and graded.

Grading

Ideally, grading should begin with a clerical procedure tabulating together the responses from all students with frequency counts. Responses can be listed in order of frequency of occurrence and then graded. Appendices D and E present such distributions. Note that unacceptable responses are infrequent. Scores may then be transposed to the responses on each of the original test papers. This procedure guarantees a high reliability in grading.

Lacking the computer technology, hand grading, as seen in Appendix D, must be relied upon. It goes remarkably easily, and does not take as much time overall as putting together an equivalent multiple choice exam.

Three methods have been used in grad-

ing, and inter-grader reliabilities measured for each.

A. Four Step Scale for Grading

+2 - A good associate, relevant, clearly demonstrating familiarity with the material.

+1 - Acceptable, but strained, ambiguous, overly general, or a reasonable guess.

0 - No response, irrelevant, indeterminate, or out of context.

-1 - Positively incorrect.

As an example of the four step system, scores would be assigned to the responses below as follows:

Binet

Intelligence test	+2
French	+1
Stephen Vincent	0
Freud's student	-1

B. Three Step Scale

The preceding system was simplified to a +2, +1, or 0 scale. The -1 score was eliminated since it seemed to measure the grader's response to the student's inadequate response. Further, relative to the number of "correct" or "acceptable" responses, few clearly incorrect rather than irrelevant responses are given, so few associates appear that get graded -1. Dropping the penalty for "wrong" answers proves not to affect the grade curve.

On the 2, 1, 0 scale, do not limit the use of 2 to indicate the "exceptionally good" response (corresponding to a -1, as somehow especially bad). A score of "2" should mean that "no one could take exception" to the associate, while a "1" indicates a response that is in the "right direction," but not adequate to "pinpoint" the stimulus term. With experience in grading, the grader will develop both speed and confidence in his or her objectivity and consistency in grading, especially if the paper graded is identified by student number, rather than by name.

C. Simplest Scale: Two Steps

The third scale is the simplest; responses are graded 1 and 0, acceptable or not.

Saeger (1975) states, "Review of the problem of item weighting in general and the magnitude of the correlations between weighted versus unweighted scoring of responses on the WAT in particular indicates that a simplified scoring system of +1 for an acceptable associate (encompassing +2 and +1) and 0 for an unacceptable associate (encompassing 0 and -1) will provide sufficient estimates of test characteristics."

D. Count Score

Those who have had considerable experience reviewing WAT scores are not surprised to find that simple counts of responses, without grading, correlate highly with scores.

PSYCHOMETRIC PROPERTIES

Reliability

Variability distributions of scores made in a class will show a broad range of performances—from those who fill in almost every blank correctly, to those whose answer sheets are less than half full.

Research on the Tennessee WAT indicates that the procedure is acceptably reliable. Various coefficients of reliability range from +.592 to +.899. See Appendix F.

Validity

The measurement of the validity of any examination of a student's achievement in a course is a matter of faith. The question of the criterion or criteria against which the validity of any one test is to be evaluated is very much open to debate, the question arising just as critically for essay, short answer, take-home, and multiple choice exams as for the WAT. Appendix F pre-

sents validity coefficients of WAT scores with the essay, short answer, and GPA. "Validity" estimates range from .447 (WAT and overall GPA) to +.773 (WAT and Short Answer Test) depending on the criterion. Most coefficients are between +.45 and +.77 and are thus well within the acceptable range.

Those who developed the test have used it in many smallish (10-20) classes in both beginning and advanced Honors Psychology groups, in which it is possible to get to know the students and their performances reasonably well. One is very seldom surprised by scores on WAT, and then it is by the score of the shy, quiet student who performs very well indeed on the Associate Test.

REFERENCES

- Attkisson, C. C., & Snyder, C. R. (1975). Student evaluation of multiple choice and word association exams. *Journal of Instructional Psychology*, 2, 9-15.
- Boneau, C. A. (1990). Psychological literacy: A first approximation. *American Psychologist*, 45, 891-900.
- Jennings, W. N. (1966). *Originality as a function of uncommon associations to common stimulus words*. Unpublished master's thesis. University of Tennessee.
- Jennings, W. N. (1968). *The assessment of instructor differences using the Word-association examination*. Unpublished doctoral dissertation. University of Tennessee.
- Rosanoff, A. J., Martin, H. E., & Rosanoff, I. R. (1918). A higher scale of mental measurement and its application to cases of insanity. *Psychological Monographs*. Princeton, N.J.
- Shelnutt, J. B. (1969). *Some characteristics of the Word-Association Test used as a measure of academic knowledge with respect to individual differences in student's patterns of responding, difficulty of stimulus items, and heterogeneity of the response distributions*. Unpublished honors thesis. University of Tennessee.
- Saeger, W. (1975). *An associative evaluation of selected interdisciplinary courses*. Unpublished doctoral dissertation. University of Tennessee.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Appendix A

List of first 194 words of (approx) 2500 on which students in a beginning General Psychology course have been examined using the Word Associate Test.

ablation	aversive stimulus	cell differentiation	consciousness
abscissa	axon hillock	central limit theorem	consistency
accountants	Babinski	central tendency	consolidation theory
Ach	balance theory	centrality index	consonance dissonance
acquired drive	balloon	cephalocaudal	constancy
acquisition	Bartlett	cerebrotonic	constant method
activation syndrome	basic anxiety	cerebrum	constitutional types
active vocabulary	Beach	chained schedule	construct validity
actuarial	behavior therapy	chance errors	consummatory act
adaptation	behaviorism	character	consummatory behavior
adaptation level	Bekesy	Charcot	
adipsia	Bekhterev	checking force	content analysis
adrenal	Bell	Chomsky	content psychology
adrenalin	Berger	choroid	context analysis
affectional drive	Berkeley	chromosome	context theory
affective congruency	beta weights	chronograph	contiguity
all or none	Binet	classical	contraction
alpha	biosocial theory	classical conditioning	control
alpha rhythm	bipolar cell	clinical interaction	convergent hierarchy
Ames	bit	closure	core and context
amnesia	black headed gull	clustering	corpal data
ampulla	blast injection	cognitive change	correlation
amygdala	blockings	cognitive dissonance	correlation asch
analyzer	Boring	cognitive maps	correspondence
animal consciousness	Bradley	cognitive theory	cortex
anticathexis	brain	cohesiveness	cortical steroid
anticipation error	brain-washing	collective unconscious	corticospinal
anxiety	Brentano	color blindness	counter conditioning
anxiety reduction	Breuer	color solid	counterbalancing
aphagia	brightness	color vision	counterconformity
aphalic index	brightness constancy	common elements	cranial nerves
approach gradient	Broadbent's model	common trait	Crespi-effect
apraxia	Broca	comparative judgement	cretinism
Aquinas	Broca's area	compromise response	criterion
area sampling	Brunswik	computers	criterion analysis
arousal	Bunsen-Roscoe law	concept formation	critical period
association test	C. Lloyd Morgan	conditional achievement	cross validation
asymptote	canalization	conditioned elements	csr
atomism	Cannon	conditioned inhibition	cultural conflict
atropine	case history	conditioned response	cumulative record
attensity	castes	cones	cumulative recorder
attention	castration complex	conflict	curare
attenuation	catatonic	conformity	curiosity
authoritarian	catharsis	confounding	dark adaptation
autism	cathexis	congruity incongruity	dat
autokinetic	Cattell	conjunction	ddd syndrome
automated teaching	cell assembly	connector neuron	De La Mettrie
autonomic	cell body	connotative meaning	decerebrate rigidity
			decibel
			decision theory

Appendix B-1
Completed Test

Psychology 251-B
Honors General Psychology
COURSE EVALUATION EXAM
Short-Answer Score 115

affective disorder affects behaviour	1	group therapy method used for treatment for abnormal people	2	personalized systems of instruction small instructor student ratio	0	socialization Sociology	1
affective disorder noticeable disorder	1	group therapy psychiatrist	1	personalized systems of instruction better environment for learning	0	socialization behavior with other organisms	2
aptitude ability to do something	2	homeostasis same level	0	phenomenology naturalism	2	statistical inference conclusion drawn from statistics	2
aptitude intelligence test	2	homeostasis stationary	1	phenomenology type of psychology	2	statistical inference human error	1
authoritarian personality someone in charge	0	insight aha!	2	prejudice discrimination	2	test battery dumb test	1
authoritarian personality instructor	0	insight thinking	2	prejudice oppression	2	test battery Binnet test	2
behavior therapy Watson	2	introvert examining self	0	psi parapsychology	2	unconditioned stimulus (USCS) automatic response	1
behavior therapy Skinner	2	introvert looking into self	0	psi ESP	2	unconditioned stimulus (USCS)	
claustrophobia fear of closed area	2	LSD drug	2	psychogenic	0	Weber's law	
claustrophobia psycotic	0	LSD hallucinate	2	psychogenic	0	Weber's law	
computer stimulation CAL	2	mesomorph fat	0	rapid eye movements (REMS) sleep	2	wire mother monkey	2
computer stimulation learning	2	mesomorph slow	0	rapid eye movements (REMS) dreams	2	wire mother monkey development	2
Down's syndrome genetic disorder	2	negative reinforcer extinction	2	reinforcing stimulus positive reward	2	species repertory common for particular species	2
Down's syndrome mongoloid	2	negative reinforcer punishment	0	reinforcing stimulus increase particular behavior	2	species repertory	
experimental design experimental setup	1	obsessive-compulsive reaction psychosis	0	Schema		V. I. schedule	
experimental design maze	1	obsessive-compulsive reaction compulsive eater	2	Schema		V. I. schedule	
frustration stress	2	paired-associate learning form of learning	2	self-regulation nervous system	2	unconditioned response natural response	2
frustration defense mechanisms	2	paired-associate learning Skinner's box	0	self-regulation automatic	2	unconditioned response automatic	1

Gestalt	extinction negative reinforcement 2	Id Frued 2	behavior repertory common behavior 0
Gestalt	extinction end of particular trait 2	Id animal 2	behavior repertory
extrasensory perception ESP 2	nonsense-syllable 7+2 0	Retroactive Interference	washoe chimp 2
extrasensory perception parasychology 2	nonsense-syllable short term memory 1	Retroactive Interference	washoe sign language 2
ethology study of behaviour in or- ganisms other than man 2	modeling way to change behavior 2	"Black Box" Stimulus Response 2	
ethology monkeys 1	modeling used in treatment of phobia 2	"Black Box" Behaviorist 2	

Appendix B-2
Completed Test

HISTORY AND SYSTEMS

In each space provided, give a different item —, word, phase, date, term, name, concept—relevantly associated in psychology with the term or name given in the space.

Score 82

Koffka Gestalt	2	Berkeley Bishop of Cloyne	2	need-press Murray	2	parallelism mind-body	2
Koffka Köhler	2	Berkeley non-existence of matter	2	need-press achievement	2	parallelism opposed dualism	0
Koffka Wertheimer	1	Berkeley Parallélíst	2	need-press affiliation	1	parallelism non-interaction	2
Koffka Pragnanz	2	Berkeley Judgement of distances	2	need-press way to satisfy need	1	parallelism contiguity but no casuality	2
formal cause Löcke	-1	secondary process Freud	2	E. C. Tolman purposive behavior	2	Ehrenfels	
formal cause		secondary process ego	2	E. C. Tolman B=f (S, A)	2	Ehrenfels	
formal cause		secondary process reality	2	E. C. Tolman cognitive map	2	Ehrenfels	
formal cause		secondary process		E. C. Tolman Holt	2	Ehrenfels	
Leipzig Wundt	2	Leibniz Monad	2	bit		Walter S. Hunter symbolic process	2
Leipzig First psychological lab	2	Leibniz petites perception	2	bit		Walter S. Hunter Delayed reactions experiment	2
Leipzig 1879	2	Leibniz apperception	2	bit		Walter S. Hunter Glass barrier	2
Leipzig Study of consciousness	2	Leibniz activity	2	bit		Walter S. Hunter memory of rat	2
Hecht		Thales		de la Mettrie mechanistic psychology	2	core-and-context	
Hecht		Thales		de la Mettrie no soul	2	core-and-context	
Hecht		Thales		de la Mettrie former clergyman	2	core-and-context	
Hecht		Thales		de la Mettrie cabanais	-1	core-and-context	
Operationism Bridgeman	2	Wittgenstein Wundt	-1	C. Lloyd Morgan Evolution	2	skepticism Hume	2
Operationism non-mentalistic	2	Wittgenstein Leipzig	-1	C. Lloyd Morgan Occam's razor	2	skepticism	
Operationism Behaviorism	2	Wittgenstein First Psychological paper	-1	C. Lloyd Morgan animal behavior	2	skepticism	
Operationism		Wittgenstein		C. Lloyd Morgan		skepticism	

Street Railways	Ecosystem	Utopia
Street Railways	Ecosystem	Utopia
Street Railways	Ecosystem	Utopia
Junior Johnson	Intuition	Gas Turbine Engine
Junior Johnson	Intuition	Gas Turbine Engine
Junior Johnson	Intuition	Gas Turbine Engine
Junior Johnson	Intuition	Gas Turbine Engine
Behavioral Sink	The Steamboat	Time Sense
Behavioral Sink	The Steamboat	Time Sense
Behavioral Sink	The Steamboat	Time Sense
Behavioral Sink	The Steamboat	Time Sense
Highway Lobby	Critical Material	Environmental Crisis
Highway Lobby	Critical Material	Environmental Crisis
Highway Lobby	Critical Material	Environmental Crisis
Highway Lobby	Critical Material	Environmental Crisis

Appendix D-1

Response distribution to the stimulus term 'cone.'
N students = 166; N possible responses = 664
Associates with F less than 5 omitted

Associate: Primary Intraverbal Response	F	RF	Associate: Primary Intraverbal Response	F	RF
rods	279	.42	3 types	11	.02
eye	254	.38	vis. acuity	9	.01
fovea	247	.37	green	9	.01
color	175	.26	night	9	.01
retina	156	.24	night vision	9	.01
vision	87	.13	in fovea	9	.01
color vision	75	.11	Young-Helmholtz	9	.01
light	42	.06	part of eye	8	.01
daylight	36	.05	yellow	8	.01
day vision	27	.04	daytime	8	.01
sight	26	.04	blind spot	8	.01
iodopsin	26	.04	optic nerve	8	.01
lt. sensitive	18	.03	cells	7	.01
3 colors	17	.03	in eyes	7	.01
receptors	16	.02	dark	7	.01
conc. in fovea	15	.02	rhodopsin	7	.01
three	14	.02	ice-cream	7	.01
day lt. vision	13	.02	Hering theory	7	.01
color blindness	13	.02	red	6	.01
yellow, green, blue	12	.02	sensitive	5	.01
blue	11	.02	blanks & lower	257	.39
perception	11	.02	frequency responses		

Appendix D-2

How concepts change

Group A

Response distributions (four associates) to items appearing on final examination of one class in General Psychology, and on "orientation" pretest of another. Number of blanks and low-frequency responses omitted. N = number of students.

1. Stimulus word: CONTROL

Pretest (N=87)			Final (N=115)		
Response	FR	F	Response	F	RF
Power (over)	31	.36	Group	58	.50
Experiment	21	.24	Experimental	54	.47
Sociological Factors	15	.17	Experi(ments, tation)	35	.30
Variables	13	.15	Independent Variable	20	.17
Self	13	.15	Variables	16	.14
Emotions	12	.14	No Special Treatment	15	.13
Stimulus (&/or) Response	12	.14	Generic	11	.09
Mind	11	.12	Baseline (basis)	9	.08
Behavior	8	.09	Validity	7	.06
Groups	7	.08	Necessary	6	.05
Reward	4	.05			
Test	4	.05			
Biological	4	.05			

2. Stimulus word: CENTRAL TENDENCY

Pretest (N=87)			Final (N=115)		
Response	F	RF	Response	F	RF
Habits, Drives, etc.	15	.17	Mode	96	.83
X, Median, Mode	12	.13	Median	96	.83
Stimulus (&/or) Response	9	.10	Mean	84	.73
Cause	7	.08	Average	14	.12
Most Likely	7	.08	Standard Deviation (S.D.)	11	.10
Physiology			Frequency Distribution	9	.08
(brain, nervous system)	6	.07	Graph	6	.05
Inward Feelings	6	.07			
Emotional (love, hate, anger)	5	.06			
Standard Deviation	4	.05			
Average	4	.05			
Behavior	4	.05			

Group B

Response distribution (four associates) to items appearing on the pretest and final examination of the same class.

1. Stimulus Word: HABITUATION

Pretest (N=68)			Final (N=60)		
Response	F	RF	Response	F	RF
Environment	10	.15	Conditioning or Learning	33	.55
Habit	5	.07	Clothes	13	.21
Continue	3	.04	Adaptation	11	.18
Stimulus	2	.03	Perception	9	.15
Desires	2	.03	Getting used to	7	.12
Need	2	.03	Accustomed	7	.12
Home	2	.03	Unaware(ness)	6	.10
Life	2	.03	Behavior	5	.08
Condition	2	.03	Adjustment	5	.08
Response	2	.03	Response	5	.08
Exhaustion	2	.03	Environment	4	.07
Reflex	2	.03	Motivation	4	.07
			Habit Formation	4	.07
			Repeated Stimuli	4	.07
			Experience	3	.05
			Babinski	3	.05

2. Stimulus Word: HOMEOSTASIS

Pretest (N=68)			Final (N=60)		
Response	F	RF	Response	F	RF
Stable	3	.04	Balance	34	.57
Internal Environment	3	.04	Autonomic	17	.28
Map	2	.03	Optimal	12	.20
Dynamic	2	.03	Equilibrium	11	.18
			Tension	9	.15
			Behavior	6	.10
			Temperature	6	.10
			Internal	4	.07

3. Stimulus Term: FUNCTIONAL AUTONOMY

Pretest (N=68)			Final (N=60)		
Response	F	RF	Response	F	RF
Movement	4	.06	Nervous System	17	.28
Muscles	2	.03	Motivation	8	.13
Self	2	.03	Miser	7	.11
Involuntary	2	.03	Homeostasis	3	.05
Nerves	2	.03			

4. Stimulus Term: ROD-CONE BREAK

Pretest (N=68)			Final (N=60)		
Response	F	RF	Response	F	RF
Eye	15	.22	Dark Adaptation	23	.38
Color	5	.07	Rods-cones	15	.25
Sight	2	.03	Rhodopsin	14	.23
Nerve	2	.03	Day-night vision	13	.22
Back of Eyeball	2	.03	Twilight vision	12	.20
			Purkinje (shift, effect or phenomenon)	11	.18
			Eyes	7	.12
			Red (lenses, vision, goggles)	7	.12
			Vision(ual)	6	.10
			Threshold	6	.10

Appendix D-3

Psychology-Zoology 472. Winter Quarter 1963-1964
Item Difficulty List - Final Exam
Maximum points per item (8 X 20) = 160
Four responses; graded 2, 1, 0, -1

Total Score on item			Total Score on item		
item	RF	Stimulus word	item	RF	Stimulus word
157	.98	peck-order	138	.86	dominance-subordination
149	.93	threat	137	.86	immobilization
148	.93	stickleback	135	.84	comparative method
146	.91	queen substance	134	.84	sage grouse
145	.91	jackdaw	134	.84	imprinting
144	.90	paleontology	134	.84	hormones of anterior pituitary
144	.90	gonadotropins	132	.83	classical conditioning
143	.89	innate	131	.82	hypothalamus
141	.88	pheromones	129	.81	linear organization
140	.88	castes	127	.79	reciprocal preening
139	.87	estrogen	126	.79	baboon harem
139	.87	waggle dance	125	.78	permanent mateship
138	.86	population pressure	120	.75	Mesozoic
138	.86	freezing	119	.74	supplementary reproductives
138	.86	instinct	118	.74	rate of response

Total Score on			Total Score on		
item	RF	Stimulus word	item	RF	Stimulus word
118	.74	hypertrophy of sex glands	76	.48	vasopression
117	.73	master gland	76	.48	ecological niche
115	.72	scouts	73	.46	MacDougall
114	.71	interspecific recognition of calls	67	.42	ecosystem
112	.70	phylogeny	67	.42	king
111	.69	LH	66	.41	vacuum activity
105	.66	operant	59	.37	cockatoo
104	.65	home range	58	.36	personal distance
104	.65	mood	54	.34	individual distance
104	.65	response repertoire	53	.33	social reflex no. 1
101	.63	S-R; S	52	.33	internal releasing mechanisms
101	.63	consummatory act	51	.32	paradigm
100	.63	habituation	50	.31	genetic control
99	.62	mobbing	48	.30	activation syndrome
97	.61	population homeostasis	37	.23	epideictic
96	.60	spacing	35	.22	ecdysion
95	.59	cortical steroid	32	.20	genetic drift
93	.58	trophallaxis	28	.18	setting operations
92	.58	sensitization	28	.18	corpora alata
88	.55	"local dialects"	28	.18	subterritory
87	.54	Tribolium confusum	25	.16	French corvids
87	.54	intra-specific fighting	24	.15	repletes
		after mutual recognition	18	.16	cycling in population
84	.53	discriminative stimulus	10	.06	statory phase
81	.51	interspecific signals	6	.04	lek
77	.48	general adaptation syndrome	2	.01	local enhancement
			0	.00	environmental continuum

Appendix D-4

The recoverability of stimulus terms

The left-hand column gives the response-distributions from which the "key words" were chosen. The right-hand column gives the response-distributions obtained when the subjects were given one blank to fill in, with an instruction; find the "key word" that links together the four words presented.

For Group I, the key-words were selected from the response-distributions given by the same class on a previous exam.

For Group II, the key-words were selected from the response-distributions given by a class other than the one seeking the key-word.

I (a) 212-2a N=74

4 Associates to Stimulus Word: Phrenology

Response	N	RF
Gall and/or Spurz.	59	.80

212-2a N=74

"Key Word" to "G. & S., bumps, localization, pseudo science."

Response	N	RF
phrenology	62	.84

localization	48	.65	repetition of S	2	.03
pseudo science	48	.65	others	10	.13
bumps	37	.50			
size	20	.27			
brain	16	.22			
trait characteristics	13	.18			

I (b) 212-2a N=74

4 Associates to Stimulus Word:
Extirpation

Response	N	RF
ablation	44	.60
removal	38	.51
study-investigate	26	.35
lesions	23	.31
localization	21	.27
brain	13	.18
drugs	6	.08
names-(Lashley, Magoun, etc.)		.08

212-2a N=74

"Key Word" to "ablation, removal,
localization, lesions."

Response	N	RF
extirpation	19	.36
brain damage	11	.31
brain	10	.13
lobotomy	7	.09
brain functions	6	.08
repetition	2	.03
others	19	.36

II (a) 212-1 N=33

4 Associates to Stimulus Word:
Summation

Response	N	RF
spatial	18	.55
temporal	18	.55
threshold	15	.45
compound stimuli	8	.24
firing of neurons	5	.15
synapse	5	.15
addition of output	3	.09
intensity	3	.09
neuron	3	.09
stimulation	2	.06

212-2b N=81

"Key Word" to "spatial, temporal,
threshold, firing of neuron."

Response	N	RF
summation	35	.43
stimulus	10	.20
others	40	.37

II (b) 212-1 N=33

4 Associates to Stimulus Term:
Myelinated Fibers

Response	N	RF
neuron	26	.70
faster transmission	20	.61
axon covering, shield	20	.61
sheath	19	.58
insulation	6	.18
nodes of Ranvier	4	.12
around axon	2	.06
not always present	2	.06

212-2b N=81

"Key Word" to "neuron, nodes of
Ranvier, insulation, faster trans."

Response	N	RF
myelin(ated)	57	.70
(covering, sheath, tissue, fibers, insulation) others	21	.30

Appendix E-1

Frequency relative to total possible
number of associates to each of four related terms

Stimulus Term	Harlow (n = 153)	Cloth mother (n = 149)	Wire mother (n = 181)	Affectional drive (n = 179)
Harlow	1.00	.28	.19	.03
cloth mother	.28	-	.33	.08
wire mother	.22	.30	-	.07
experiment	.06	.06	.07	.00
Sci. American	.06	.01	.00	.00
behavior	.05	.00	.00	.01
imprinting	.04	.08	.08	.07
instinct	.04	.03	.05	.09
rats	.04	.00	.01	.02
motivation	.03	.00	.03	.20
monkey baby	.03	.03	.04	.00
wire	.03	.06	.00	.00
cloth	.03	.00	.04	.00
maternal	.02	.03	.00	.14
surrogate	.02	.03	.06	.00
maternal instinct	.02	.02	.03	.03
feeding	.01	.03	.06	.00
Hebb	.01	.03	.02	.01
love	.00	.03	.02	.28
sex	.00	.00	.00	.12
security	.00	.09	.02	.02
emotion	.00	.00	.03	.08
innate	.00	.00	.00	.10
fear	.00	.01	.06	.02
affection	.00	.04	.04	.00
hypothalamus	.00	.00	.00	.07
protection	.00	.03	.03	.00

Each stimulus term was given to approximately 1/4 of the same class on a final examination. Four responses were possible.

Appendix E-2

The probabilistic relationship of
primary intraverbals in a group

133 students took this WAT; all stimulus terms were presented four (4) times. The terms in squares were presented as stimuli in the WAT. The encircled terms appeared as responses (see page 114). Arrows connecting S and R give relative frequency of occurrences of each response relative to the maximum possible number (532). The careful reader will note that one response that occurred relatively frequently is not appropriate ("wrong"). Investigation showed that this word appeared in a paragraph of the text book that included the stimulus term and two or three "correct" associates. (Draw your own conclusions.)

Appendix F-1

Some reliability and validity measures of the WAT and other tests

Introductory Psychology, Honors Section *

I. Reliability

A. Split half

1. First half-second half: $r = +.680$ 2. Odd-Even: $r = +.686$

3. Ranked difficulty of items:

 $r = +.683$

B. Kuder-Richardson formula 20

1. Original test (36 items): $r = +.724$ 2. Original test (1st 20 items only):
 $r = +.592$ 3. Retest (20 items): $r = +.736$

C. Test-retest (20 items)

1. Students' scores: $r = +.624$ 2. item difficulty: $r = +.899$

II. Validity

A. Rank-order correlation between test part scores

1. Essay and short answer: $+ .804$ 2. Association and short answer:
 $+ .733$ 3. Association and essay: $+ .602$

B. Correlation between GPA and scores of:

1. Association part: $r = +.447$ 2. Entire test: $r = +.382$ 3. Short answer part: $r = +.265$ 4. Essay part: $r = +.254$

C. Correlation between position on Dean's List and rank of score on:

1. Association part: $r = +.678$ 2. Entire test: $r = +.677$ 3. Short answer part: $r = +.609$ 4. Essay part: $r = +.409$ * *Not an exhaustive set.***Appendix F-2**Some correlations on
five examinations

N = No. students n = No. st terms

4 resp; 2, 1, 0, -1 grading

1. General Psych., weekly quiz

N = 133n = 25

WAT: Mean score 137.0 SD = 23.3

% of max score possible = 68.5

GPA: Mean 2.30 SD = 0.54

int. consist (KR) $r = 0.816$ WAT with GPA $r = 0.356$ WAT with M.C. $r = 0.459$ M.C. with GPA $r = 0.359$

2. General Psych., Honors Final Exam

N = 25

n = 36

M = 180.2

SD = 28.2

% of max score = 62.6

M = 3.44

SD = 0.28

(KR)

 $r = 0.724$

WAT with GPA

 $\rho = 0.447$

WAT with short ans.

 $\rho = 0.773$

WAT with essay

 $\rho = 0.602$

Short ans. with GPA

 $\rho = 0.265$

Essay with GPA

 $\rho = 0.254$

3. Comparative Psych., Upper Div.

N = 20

n = 81

WAT: Mean score 388.1 SD = 69.3

% of max score possible = 59.9

GPA: Mean score 2.52 SD = 0.44

WAT with GPA $\rho = 0.647$

4. History and Systems, Upper Div.

N = 28

n = 49

M = 218.39

SD = 78.1

% of max score = 65.8

M = 2.73

SD = 0.87

WAT with GPA

 $\rho = 0.455$

Appendix G-1
Research on Honors Comprehensive Examination

Research on Undergraduate Honors Comprehensive examinations given at the end of the first year of the Honors program. If a student does not pass at that time he or she is examined the following year. The Comprehensive Examination has ten subsections, of which History and Systems is one. Different samples of terms are prepared for each administration. The Comprehensive Exam was taken by Advanced Honors Students at the end of the Junior Year, which had been devoted to intensive readings, with weekly seminars.

The same examination was given experimentally to graduate students in Psychology at the Master's level to replace a departmentally prepared ten-section multiple-choice examination. It was replaced by the appropriate Graduate Record Examination, which placed no demands on faculty time.

Honors Comprehensive Examination
Means and Standard Deviation
(% maximum possible score)

Year	Form; # Items	Section:	Sensation and Perception	History and Systems	Personality	Physiological	Symbolic Processes	Learning	Motivation	Psychometrics	Social	Mean (SD)
1976A	A	M	56.7	67.4	73.6	51.2	52.9	63.0	73.1	55.3	74.6	64.0
N 10	216	SD	25.6	19.8	22.7	21.4	19.8	16.6	21.2	23.9	22.8	(9.4)
1978	A	M	48.5	60.5	69.9	70.1	48.4	72.4	63.8	41.2	50.2	58.3
N (10)	288	SD	22.7	17.2	17.9	12.1	21.8	15.8	16.5	18.9	17.5	(11.5)
1975	B	M	44.6	65.2	73.6	65.2	48.1	68.5	72.9	55.5	58.7	61.4
N 11	576	SD	19.6	17.9	14.8	18.2	10.5	19.1	15.4	16.3	15.4	(10.3)
1977	B	M	59.9	71.2	80.2	67.8	47.1	72.1	66.7	51.4	56.0	63.6
N 6	288	SD	15.7	4.4	7.2	12.5	12.8	11.6	13.1	11.7	8.7	(10.7)
1979	C		37.3	42.4	67.6	58.5	51.8	57.1	64.3	45.2	49.7	52.6
N 8	288		14.1	20.1	16.1	12.7	17.6	7.9	15.2	17.1	18.5	(10.1)
All	M		52.4	66.2	74.3	65.6	49.1	69.0	69.1	64.7	59.9	
45	SD		20.9	19.3	15.7	16.1	16.2	15.8	16.5	17.7	16.1	

Internal Reliability

1) Undergraduate Comprehensive Exam

# Items	Year	Pearson's R odd-even										All Exams	
		Sensation and Perception	History and Systems	Personality	Physiological	Symbolic Processes	Learning	Motivation	Psychomatics	Social		Alpha Coefficient	
216	10	1976	.87	.91	.89	.83	.75	.88	.75	.93	.82	.96	.82
288	10	1978	.96	.83	.71	.86	.93	.91	.82	.63	.76	.96	.80
576	11	1975	.86	.49	.39	.69	.42	.76	.84	.77	.65	.94	.92
288	6	1977	.88	.23	.50	.46	.94	.44	.89	.64	.61	.82	.87
352	8	1979	.67	.94	.15	.72	.89	.13	.81	.78	.92	.91	.92
		M	.85	.68	.53	.71	.79	.62	.82	.75	.75	.90	.88
		SD	.11	.31	.29	.16	.22	.33	.05	.12	.13	.07	.06

N = 45

$\bar{x}$.722

SD .21

2) Scores on a Graduate Examination;
Ethological Psychology

	actual data		interchanged item scores		interchanged subject scores	
α items	$\bar{x}$ = 6.32	SD = .99 α = .91	$\bar{x}$ = 6.32	SD = .90 α = .88	$\bar{x}$ = 6.32	SD = .99 α = -.50
α students	$\bar{x}$ = 6.32	SD = .132 α = .65	$\bar{x}$ = 1.32	SD = 1.32 α = .55	$\bar{x}$ = 6.32	SD = .41 α = .46

N = 14

items = 19; 4 responses
2, 1, 0 scoring

α measures not only the *item* reliability, but also the consistency of student's scores.

Validity or
Self-Fulfilling Prophecy?
WAT Comprehensive Examination
for Undergraduate Honors Students

Students enter the program at the beginning of their Junior Year. Through the year, they are expected to achieve mastery over the nine fields of psychology, largely by use of a reading list in guiding independent study. In May of the Junior Year, they take the WAT Comprehensive Examination. If they do not pass the Examination at this time, they may retake appropriate sections of the Exam in the following Fall Quarter, or they may drop out of the program.

A Passing Score is 60.0 overall, with no section below 50.0 (depends on section).

Mean score, over all sections: (% maximum possible score), on *first* examination at end of Junior Year.

M = 61.4

SD = 14.6

N = 45

Score	Bottom Third 25.6-58.5	Middle Third 58.6-68.5	Top Third 68.6-84.6	All	X ²	b
Admitted to Graduate School	2	1	8	11	7.18	<.02
Unsuccessful Application to Grad. School	0	1	2	3		
Admitted, Law or Medical	0	2	1	3		
Graduated; no Graduate School	0	6	0	6	12	<.01
Dropped out	7	1	0	8		
Now Completing B. S.	6	4	4	14	10.7	<.01
All	15	15	15			

Appendix H

Interscorer Reliability of Scoring

A. Reliability of Exam Grade

class:
2518-Fall 1979

14 students
47 items
2 graders

2 responses; 2-1-0 grading

	No. Responses (94 max.)	Grader A	Grader B
M =	43.3	53.4	57.4
SD=	21.5	24.0	27.2

t = significant at the
.01 level

Pearson Product				Spearman Rank Order			
	Nr	A	B		Nr	A	B
Nr	—	.82	.86	Nr	—	.82	.91
A	—	—	.99	A	—	—	.97

If the score 2 is allotted to all 1's the inter-grader Pearson r falls to .97.

Conclusions: a. Although there is a small constant difference between the two graders, it does not affect the individual's standing in the group.

b. A grade adequate for many purposes (e. g. on a Diagnostic Exam) may be obtained by counting blanks, irrespective of the quality of the responses given.

Integrader Agreement on
Scoring of Responses (2, 1, 0)
2518 Pre exam

	1975 (AD)		1976 (AC)		1977 (AD)		1978 (AD)		All
	N	RF	NRF	RF	N	RF	N	RF	
Same grade	453		690		470		492		2105
		.68		.72		.77		.81	.74
1 point diff.	176		217		117		103		613
		.26		.23		.19		.17	.22
2 point diff.	37		47		26		11		121
		.06		.05		.04		.02	.04
Total Resp.	666	(.58)	954	(.60)	613	(.39)	606	(.46)	2839
Blanks	474		630		971		710		2785
Total Pos. R's	1140		1584		1584		1316		5624

Integrader Agreement
2, 1, 0 per item responses
2518 Exams

1. Integrader agreement on responses seems independent of over-all performances on the test.
2. A significant value of χ^2 (P.S. .001) suggests that there is progressive increase in agreement perhaps with restatement of the criterion.

B. Reliability of score of individual responses

Measures of Inter-observer (scores). Agreement between two scores of one WAT exam

Number possible responses = 1316
Number of blanks = 710
Number of responses = 606

Measure of Agreement	Score	2	1	0	2+1+0	2 or 1	2 or 1; 0	1 or 0	2; 1, 0
	% Agreement	47.7	13.0	20.5	81.2	88.0	90.4	78.9	89.1
Efficiency %		81.2	43.4	68.1					
Kappa		.271	.088	.152	.691	.396	.780	.249	.779
,Ø		.779	.501	.759					
d' A/B		2.48	1.50	2.75					
B/A		2.47	1.55	2.60					

Conclusion: The score "1" may expedite grading, but in arriving at a grade, answers scored "1" may be counted as "2."

TABLE 1
Correlations based on selected students.
Data on graduate students on whom data were available
Intercorrelation of measures. (Max N)

VARIABLE	1	2	3	4	5	6	7	8	9	10
1	—									
2	.634(57)	—								
3	.524(56)	.610(62)	—							
4	.273(46)	-.004(46)	.188(45)	—						
5	-.072(53)	.043(53)	.297(46)	.292(46)	—					
6	-.093	-.271(53)	.093(53)	.202(46)	.292(46)	—				
7	-.099(53)	.381(53)	.043(53)	.297(46)	.728(54)	.859(54)	—			
8	.000(51)	.144(52)	.071(51)	-.130(43)	-.141(51)	-.171(51)	-.193(51)	—		
9	.004(47)	.087(47)	.080(47)	-.132(40)	.027(46)	-.347(46)	-.225(46)	.307(45)	—	
10	0.58(44)	-.012(45)	.056(45)	.024(37)	.369(44)	.250(44)	.361(44)	.271(44)	.013(41)	—

N. B. (With each correlation is the number of students on whom data were available in the variables correlated. The students were selected on the basis of the year they entered graduate study, and had completed or dropped out graduate school)

- 1: Final Degree Attained (rbw)

2: Word-Association Comprehensive Examination

3: Multiple-Choice Comprehensive Examination

4: Graduate Record Examination-Advanced (Psychology Section)

5: Graduate Record Examination-Verbal
- 6: Graduate Record Examination-Quantitative

7: Graduate Record Examination Combined (V + Q)

8: Incoming Grade-Point Average

9: Mean Value of Letters of Reference

10: Mean Faculty Rating of Applicant's Folder

TABLE 2
Correlations for selected students with information on all variables.
Intercorrelation of measures
(N=33)

VARIABLE	1	2	3	4	5	6	7	8	9	10
1	—									
2	.621	—								
3	.482	.537	—							
4	.191	.104	.335	—						
5	-.182	-.335	-.147	.154	—					
6	-.038	-.416	.010	.278	.335	—				
7	-.155	-.464	-.067	.275	.744	.879	—			
8	.166	.221	-.025	-.151	-.107	-.251	-.232	—		
9	-.084	.051	-.084	-1.63	.000	-.363	-2.57	.304	—	
10	.044	.043	.000	.043	.373	.220	.345	.204	-.038	—

N. B. (Of the students reported on Table 2, on only 33 were scores on all ten items available.)

- 1: Final Degree Attained (tbw)

2: Word-Association Comprehensive Examination

3: Multiple-Choice Comprehensive Examination

4: Graduate Record Examination-Advanced (Psychology Section)

5: Graduate Record Examination-Verbal
- 6: Graduate Record Examination-Quantitative

7: Graduate Record Examination Combined (V + Q)

8: Incoming Grade-Point Average

9: Mean Value of Letters of Reference

10: Mean Faculty Rating of Applicant's Folder

TABLE 3

Factor analysis of intercorrelation matrix.
Varimax factor loadings for selected students
with complete data (N=33)

FACTOR				
Variable	I	II	III	h?
Final Deg Att.	.714	-.004	-.004	.505
WAT Comps	.870	-.291	-.124	.856
MC Comps	.687	.210	-.050	.518
GRE-ADV	.287	.421	.126	.275
GRE-V	-.237	.199	.509	.355
Letters of Reference	-.069	-.475	-.021	.230
Incoming GPA	.151	-.519	.174	.322
Mean Fac Rating	.093	-.093	.783	.639