

TEST ITEM FILE

Psychological Testing

An Introduction

SECOND EDITION

George Domino

University of Arizona



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

32 Avenue of the Americas, New York, NY 10013-2473, USA

© Cambridge University Press 2007

This publication is in copyright. Users may print out all or parts of the work for private or research or educational use in conjunction with the textbook Psychological Testing: An Introduction by George and Marla L. Domino (Cambridge University Press, 2006), but may not reproduce it for any other purpose, nor disseminate it beyond an immediate classroom context, without the written permission of Cambridge University Press.

Contents

Foreword	page vii
1 The Nature of Tests	1
2 Test Construction, Administration, and Interpretation	6
3 Reliability and Validity	15
4 Personality	23
5 Cognition	35
6 Attitudes, Values, and Interests	48
7 Psychopathology	60
8 Normal Positive Functioning	71
9 Special Children	80
10 Older Persons	93
11 Testing in a Cross-Cultural Context	99
12 Disability and Rehabilitation	110
13 Testing in the Schools	120
14 Occupational Settings	132
15 Clinical and Forensic Settings	144
16 The Issue of Faking	154
17 The Role of Computers	166
18 Testing Behavior and Environments	174
19 The History of Psychological Testing	186

Psychological Testing: An Introduction, 2e

George Domino
Department of Psychology, University of Arizona

Foreword

This instructor's manual contains multiple-choice items for Psychological Testing: An Introduction. With a few exceptions, the items are listed in order as they occur in the textbook, and the correct answer is marked for each.

In preparing an exam using these items, two cautions need to be made:

- A) some of the items may well provide clues to each other, particularly if they are presented in the same sequence as in this manual.
- B) when taken out of their context, a few items may need to be revised in the stem portion to make clear to the student what is being asked.

1 The Nature of Tests

1. In defining a test, the term objectivity refers to:
 - *A) objective criteria used in scoring and interpretation
 - B) an ideal rarely achieved
 - C) the degree of competence shown by the examiner
 - D) uniformity of procedure
2. The term objectivity refers to:
 - A) uniformity of procedure
 - B) the use of multiple choice items
 - *C) objective rather than subjective criteria
 - D) assessment rather than testing
3. The metaphor of a test as an experiment focuses on:
 - *A) keeping all extraneous variables in check
 - B) the subject rather than the experimenter
 - C) sampling accurately among potential subjects
 - D) the ease of testing many subjects at one sitting
4. Standardization means:
 - A) objectivity of scoring
 - B) the use of standard scores
 - C) experimental control over extraneous variables
 - *D) uniformity of procedure
5. Thinking of a test as an interview focuses on:
 - A) the standardization that interviews lack
 - B) the fact that interviews cannot be faked
 - *C) the necessity and objectivity of the interview
 - D) the advantages of a one-on-one approach
6. Domino uses the term “psychological test”
 - A) in a very narrow sense to refer to personality tests only
 - B) as a synonym for intelligence tests only
 - *C) in a very broad sense
 - D) to refer to any type of measurement
7. A battery consists of:
 - A) tests used only for clinical assessment
 - B) tests that are not standardized
 - C) standardized tests like the Stanford-Binet
 - *D) a group of tests

The Nature of Tests

8. The Oakland and Hu (1991) survey of who uses tests found that:
 - A) only psychologists used tests
 - B) locally developed tests were the most common
 - C) testing is strictly an American phenomenon
 - *D) a variety of professional groups use tests
9. Placing a student in introductory rather than advanced German on the basis of test scores involves:
 - *A) classification
 - B) self-understanding
 - C) program evaluation
 - D) scientific inquiry
10. Departing from standard administrative procedures in giving a test:
 - *A) may influence the obtained results
 - B) does not influence the obtained results
 - C) always results in incorrect findings
 - D) is entirely up to the examiner
11. An example of a situational variable that can influence test results:
 - A) an aloof examiner
 - *B) administering the test right after lunch
 - C) how attractive the examinee is
 - D) instructions to fake the test
12. Which is most correct? The race of the examiner:
 - *A) is not as powerful an influence as one might think
 - B) plays a major role in the obtained test results
 - C) must match the race of the subject for valid results
 - D) is important only with male examiners
13. Departures from standard administration procedures are more likely to affect:
 - A) college students
 - B) normal adults
 - *C) children
 - D) none of these
14. In decision making, such as college admissions most psychologists would argue that:
 - A) only objective test results be used
 - *B) decisions be based on empirical evidence
 - C) test scores are typically not useful
 - D) decisions be based on individual merits
15. The items in a test must directly cover the behavior of interest.
 - A) True
 - *B) False
 - C) only if the test is standardized
 - D) only if a classification decision is to be made
16. A proprietary test is one that is:
 - *A) commercially published
 - B) objective and standardized
 - C) used for scientific inquiry only
 - D) aimed at self-understanding
17. Major tests like the Stanford-Binet:
 - *A) are available for purchase by qualified users
 - B) are available for purchase by anyone
 - C) are not proprietary tests
 - D) are all criterion-referenced

18. A specimen set refers to:
 - A) physiological tests such as those used for drug usage
 - B) a free sample supplied by a test publisher
 - *C) a sample that can be obtained for a fee
 - D) a sample that can be checked out from the library
19. Tests that are not commercially published:
 - A) can be photocopied and reproduced without permission
 - B) are usually not valid
 - C) are typically reviewed in the MMY
 - *D) can usually be obtained from the author
20. The Mental Measurements Yearbook:
 - *A) provides reviews and other information for many commercially available tests
 - B) is published yearly as its name indicates
 - C) gives specimen questions for most intelligence tests
 - D) covers only tests published in a particular year
21. By definition, a group test:
 - A) cannot be an individual test
 - B) can sometimes be administered to small groups
 - *C) can usually if not always be administered to one person
 - D) is longer and more standardized than an individual test
22. If there is a time limit on a power test, the time limit is set so that about _____ % of the applicants can attempt all the items.
 - A) 10
 - B) 50
 - *C) 90
 - D) 100
23. As far as medium is concerned, most tests are _____ tests.
 - A) performance
 - B) invasive
 - C) power
 - *D) paper-and-pencil
24. The term "screening" usually refers to:
 - A) tests used for placement purposes
 - B) tests related to a legal standard
 - *C) a rapid and somewhat rough procedure
 - D) the issue of certification
25. Norm reference refers to:
 - *A) the interpretation of scores using norms
 - B) the studies by Norm and his colleagues
 - C) an examination like a driver's license test
 - D) only norms that are expressed in z scores
26. The same test
 - *A) could be both norm-referenced and criterion-referenced
 - B) cannot be both norm-referenced and criterion-referenced
 - C) has to be criterion-referenced if it involves placement
 - D) must be norm-referenced if used for screening
27. if the scores on a scale sum to a constant then the test is:
 - A) a power test
 - B) an invasive procedure
 - C) based on rankings
 - *D) an ipsative measure

The Nature of Tests

28. Impressionistic and psychometric interpretations of a test score:
- A) cannot be combined
 - *B) can sometimes be combined
 - C) apply only to power tests
 - D) are applicable only to ipsative measurement
29. A personality test would ordinarily be considered:
- A) a screening procedure
 - B) an ipsative measure
 - *C) a typical performance test
 - D) a proprietary test
30. A free response test might be illustrated by:
- A) an ipsative measure that involves ranking
 - B) a personality test used for placement purposes
 - *C) the Rorschach Inkblot technique
 - D) an interval scale of multiple choice items
31. The numbers on basketball players' uniforms represent a(n) _____ scale.
- *A) nominal
 - B) ordinal
 - C) interval
 - D) ratio
32. Ordinal scales
- *A) involve ranks
 - B) have a true zero
 - C) have equal intervals
 - D) have an arbitrary zero
33. Strictly speaking, a typical true-false personality scale is an example of a(n) _____ scale.
- A) nominal
 - *B) ordinal
 - C) ratio
 - D) interval
34. A typical multiple-choice vocabulary test
- A) uses items all of equal difficulty
 - B) gives fewer points to easier items
 - *C) assumes that we have an interval scale
 - D) is, for scoring purposes, treated as a nominal scale
35. Temperature as measured by the Fahrenheit scale is a good example of a(n) _____ scale.
- A) ratio
 - *B) interval
 - C) ordinal
 - D) nominal
36. A psychologist who claims that Test X can predict a behavior with 100% accuracy, most likely is breaking the ethical principle of:
- A) competence
 - *B) integrity
 - C) respect of privacy
 - D) concern for others' welfare
37. The ethical standards do not cover:
- A) the use of obsolete tests
 - B) familiarity with psychometric issues
 - *C) the market value of tests
 - D) the security of tests

38. Typically, informed consent means that the client
 - A) must have a clear understanding of the technical information
 - *B) must be told in general terms what the purpose and use of a test are
 - C) must have completed a basic course in psychological testing
 - D) all of the above
39. In considering ethical issues, three areas are most important:
 - A) qualifications of the examiner, privacy, and feedback
 - *B) informed consent, confidentiality, and privacy
 - C) feedback, rapport, and informed consent
 - D) validity, reliability, and norms
40. Ethical standards that govern tests were developed by:
 - A) individual psychologists
 - *B) professional organizations
 - C) the American Psychological Association
 - D) the courts as a result of civil suits
41. According to the APA scheme, level A tests
 - *A) require minimal training
 - B) are available only to Ph.D. level clinical psychologists
 - C) are given excellent reviews in the MMY
 - D) can only be scored from an ipsative framework
42. The Social Sciences Citation Index
 - *A) lists references cited in various journal articles
 - B) gives detailed reviews of most major tests
 - C) tells which tests are cited as most popular
 - D) all of the above
43. Tests in Microfiche
 - *A) can be useful to obtain information about tests
 - B) are only available from Educational Testing Service
 - C) are simply a computerized version of the MMY
 - D) is a journal that routinely publishes test reviews

2 Test Construction, Administration, and Interpretation

1. In constructing a test, the first step is to
 - *A) identify a need
 - B) develop a theoretical framework
 - C) construct a pool of items
 - D) generate a table of specifications
2. Domino argues that
 - *A) most tests are influenced by theory
 - B) most tests lack a specific theoretical framework
 - C) a test developer is better off without a theory
 - D) theoretical frameworks impose limitations on tests
3. A table of specifications refers to
 - A) how long the test will be
 - *B) a blueprint of the structure of the finished test
 - C) the practical decisions that need to be made
 - D) the degree to which theory dictates content
4. In general, an initial pool of items needs to be _____ as large as the number of items actually needed.
 - A) twice
 - *B) four or five times
 - C) ten times
 - D) depends upon the nature of the test
5. Commercially developed tests often
 - A) start out with a small pool of items
 - B) have a weak or nonexistent theory
 - *C) involve a number of experts
 - D) do not use a table of specifications
6. Professor Smith who teaches an American history course, is putting together a 100 item exam. You would advise him to initially write
 - A) 110 items
 - B) 200 items
 - *C) 400 to 500 items
 - D) only the items he will actually use
7. Administering a preliminary test form to a group of subjects:
 - A) content analysis
 - B) table of specifications
 - C) standardization
 - *D) pilot testing

8. I am developing a test to measure attitudes toward suicide. I have a pool of some 3000 potential items. A good technique at this point would be:
- *A) content analysis
 - B) pilot testing
 - C) item analysis
 - D) standardization
9. Reliability basically means:
- A) the test measures accurately
 - *B) the test measures consistently
 - C) the test scores are standardized
 - D) norms are available
10. In regards to sample size of test norms:
- A) raw scores are most meaningful
 - B) larger is always better
 - C) groups differing in age should always be used
 - *D) representativeness is important
11. A test developer should always
- A) use derived scores rather than raw scores
 - B) use normative samples defined by income and occupation
 - *C) clearly define the nature of the normative sample
 - D) use children as normative samples
12. As a general rule
- *A) the shorter the test the less reliable and valid it is
 - B) the shorter the test the more reliable it is
 - C) tests revised by factor analysis are more valid
 - D) multivariate tests are better
13. A multivariate test is one that
- A) yields several types of scores
 - B) has different forms for different age ranges
 - *C) is composed of many scales
 - D) can be analyzed statistically
14. An essay exam represents a
- *A) constructed response item
 - B) selected response item
 - C) distractor
 - D) keyed response
15. A selected response item is best illustrated by:
- A) vignettes
 - *B) multiple choice items
 - C) essay exams
 - D) multivariate tests
16. This part of a multiple choice question is called the:
- A) response option
 - B) keyed response
 - *C) stem
 - D) distractor
17. Distractors should be
- *A) equally attractive to the test taker
 - B) clearly correct or incorrect
 - C) substantially shorter than the stem
 - D) more confusing to those who know the correct answer

18. Multiple choice items have been criticized because
 - *A) they promote rote learning rather than problem solving
 - B) poorer students often do better on them
 - C) they can be scored quickly
 - D) they are less valid than constructed response tasks
19. From a psychometric point of view factual true-false items:
 - *A) are not very useful
 - B) are extremely useful
 - C) cannot be easily scored
 - D) involve too many options
20. An odd-man-out type of item
 - *A) refers to a set where one component does not belong
 - B) a sequence of items, with the last item blank
 - C) a matching set of items of unequal length
 - D) an extra item that cannot be scored
21. One problem with matching items is that
 - A) the distractors are too obvious
 - B) they can be used only with analogies
 - C) the keyed response cannot be determined
 - *D) the components are not independent
22. A variant of the completion item:
 - A) odd-man-out
 - B) vignettes
 - *C) fill in the blank
 - D) analogies
23. From a psychometric point of view, the best items are _____ items.
 - *A) multiple choice
 - B) constructed response
 - C) ipsative
 - D) rearrangement
24. To statistically analyze the results of a test, we need
 - A) as few individual differences as possible
 - B) a very small range of scores
 - *C) variation in raw scores
 - D) for the raw scores to be changed to derived scores
25. The spiral omnibus format involves test items
 - *A) listed from easy to difficult, with several such sequences
 - B) listed from easy to difficult in one sequence only
 - C) that are not counted for scoring purposes
 - D) that are criterion-keyed
26. Filler items are used
 - A) to make the test longer
 - B) to offset fatigue
 - *C) to disguise the purpose of the test
 - D) as distractors
27. "Authentic" measurement refers to:
 - A) the assessment of faking on a test
 - B) whether a test has been copyrighted or not
 - C) whether a test is the original or a revised form
 - *D) performance that can be measured directly

28. Observing whether a person can do the back stroke in a swimming pool:
- A) spiral omnibus
 - B) selected response
 - *C) authentic measurement
 - D) odd-man-out
29. The method of fiat refers to
- *A) a decision based on authority
 - B) Flexibility In A Test
 - C) criterion-keyed tests only
 - D) scales to measure depression
30. Criterion-keyed tests are developed based on
- A) the test author's authority and expertise
 - *B) the statistical relationship of an item to a criterion
 - C) the level of item difficulty found
 - D) factor analysis
31. The major challenge of criterion-keyed tests is the choice of
- A) keyed items
 - *B) criterion used
 - C) homogeneous items
 - D) a spiral omnibus format
32. Criterion-keyed scales are typically
- A) homogeneous
 - *B) heterogeneous
 - C) made up of multiple choice items
 - D) shorter than those developed by fiat
33. A potential problem with criterion-key methodology is that retained items
- A) do not work well, even when the theory is explicit
 - B) have high reliability but low validity
 - *C) will work well, but the theoretical rationale won't be clear
 - D) have low reliability and high validity
34. Criterion-keyed scales typically use criteria that are
- *A) complex
 - B) clearly defined
 - C) theoretically based
 - D) weak
35. In the example of the Rehfisch scale of personal rigidity, the criterion against which the scale was developed was
- *A) staff ratings of rigidity
 - B) self ratings of rigidity
 - C) another scale of rigidity
 - D) an experimental task
36. Factor analysis as a way of test construction assumes that
- A) scales within an instrument correlate highly with each other
 - B) all items that make up a scale do not correlate with each other
 - C) cross validation is not necessary
 - *D) scales should be univariate and independent
37. A factor loading reflects the correlation between
- A) two different items on the same test
 - *B) an item and a theoretical dimension
 - C) an item and the total score on a test
 - D) two theoretical dimensions

38. The major limitation of percentile scores is that they
- A) are difficult to compute and comprehend
 - *B) represent an ordinal scale rather than an interval scale
 - C) represent a ratio scale but the intervals are unequal
 - D) cannot be translated into z scores
39. Cecilia obtains a raw score of 85 on a test where the mean is 70 and the SD is 10. Her raw score equals a z score of
- A) -1.5
 - B) $+2.8$
 - *C) $+1.5$
 - D) $+3.5$
40. Arthur obtained a raw score of 120 on a history test where the class average was 116 and a SD of 7. He also obtained a raw score of 52 on a geography test where the mean was 48 and the SD was 5. We can conclude that
- A) he did better on the history exam
 - *B) he did better on the geography exam
 - C) he did equally well on both exams
 - D) we need to know what the range of scores was
41. On the above item, Arthur's z score on the history test was _____, and on the geography test was _____.
- A) $-.57$; $-.80$
 - *B) $+.57$; $+.80$
 - C) $-.57$; $+.80$
 - D) $+.57$; $-.80$
42. If we now change Arthur's raw score on the history test to a T score, his T score will be
- A) 44.3
 - *B) 56
 - C) -56
 - D) 107
43. Marilyn took the SAT and reports that her verbal score was 686. We would conclude that
- A) she is very bright, at least in verbal skills
 - B) she will probably do quite well in college
 - *C) her report is incorrect
 - D) her equivalent T score is 100
44. The mean of stanines is
- A) 2
 - *B) 5
 - C) 50
 - D) 100
45. A score of 790 on the SAT would most probably equal a stanine of
- A) 1
 - B) 3
 - C) 7
 - *D) 9
46. Stanines differ from z scores and T scores in that
- *A) they force the raw scores into a normal distribution
 - B) they can be computed only when the raw scores are normally distributed
 - C) they require complex calculations
 - D) their mean is always 100

47. The difficulty of an item is defined as
- A) the percentage of persons who answer incorrectly
 - *B) the percentage of persons who answer correctly
 - C) the actual number of people who know the right answer
 - D) the actual number of people who don't know the right answer
48. On a personality test, item difficulty is defined as
- A) the most popular response option that is endorsed
 - *B) the percentage who endorse the keyed response
 - C) how clear and understandable an item is
 - D) the percentage who do not endorse the keyed response
49. Professor Hindsight puts together a final exam of 180 items for his large American History course. He would like the test to be of optimum difficulty. He therefore should expect an average of
- A) 18
 - B) 70
 - *C) 90
 - D) 120
50. Professor Hindsight also develops a 180 item test to identify the top 10% of history majors, for possible scholarship awards. On this test he should expect an average of
- *A) 18
 - B) 70
 - C) 90
 - D) 120
51. Which of these z scores might represent a very difficult item?
- A) -2.06
 - B) $-.52$
 - C) $+.20$
 - *D) $+2.33$
52. A z score of $+1.00$ would equal a delta score of
- A) 13
 - *B) 17
 - C) 21
 - D) 28
53. Delta scores have a mean of _____ and a SD of _____ .
- A) 50; 10
 - B) 8; 2
 - *C) 13; 4
 - D) 100; 16
54. A test whose item difficulty distribution follows a normal curve is said to have
- A) low fidelity but large bandwidth
 - *B) high fidelity but little bandwidth
 - C) low fidelity and little bandwidth
 - D) high fidelity and large bandwidth
55. Guessing _____ the p value.
- A) decreases
 - *B) increases
 - C) invalidates
 - D) is unrelated to

56. To minimize the effects of guessing we can use
- A) delta scores to assess the difficulty of items
 - *B) instructions that specify what to do
 - C) instructions that forbid guessing
 - D) items whose difficulty level is normally distributed
57. Item discrimination refers to the ability of a test item to identify
- *A) those who are higher from those who are lower on the tested dimension
 - B) those who obtain a perfect score from all others
 - C) those who score above the median versus below the median
 - D) those who are being discriminated against
58. In computing item discrimination indices, the best strategy from a statistical point of view is to compare
- A) those above the median versus those below
 - B) "A" students versus failing students
 - *C) upper 27% versus the lower 27%
 - D) those with delta scores of above 50 versus below 50
59. The index of discrimination is typically expressed as a
- A) z score
 - *B) percentage difference
 - C) percentile
 - D) score with mean of 500 and SD of 100
60. In computing the index of discrimination, the extreme groups are identified on the basis of
- *A) total test score and/or outside criterion
 - B) total test score only
 - C) outside criterion only
 - D) none of these
61. Given the same pool of items, Dr. A attempts to develop a scale of depression through factor analysis, while Dr. B develops it through empiricism. They will
- A) come up with essentially the same scale
 - *B) come up with different scales
 - C) both fail to develop useful scales
 - D) both develop scales with equal means and SDs
62. Item Response Theory assumes that
- A) a test score is the sum of a true score and an error component
 - B) responding to an item represents a reflex
 - *C) performance on a test is a function of an unobservable proficiency variable
 - D) both low scorers and high scorers can do equally well if given the opportunity
63. A stratified sample is typically chosen on the basis of
- A) high z scores
 - *B) census information
 - C) convenience
 - D) theoretical reasons
64. In general, norms based upon samples of convenience
- A) are totally worthless
 - B) are very misleading
 - C) are acceptable if expressed as T scores
 - *D) can be useful

65. Adult age norms would be most useful on a test to measure
- A) vocabulary level
 - *B) attitudes towards suicide
 - C) physical height
 - D) interest in oceanography
66. To compute an expectancy table we need to have
- A) one set of scores that are independent of the other variable
 - B) all the scores expressed as percentiles
 - *C) two sets of scores that show some correlation
 - D) both delta scores and z scores
67. Local norms
- A) are always better than national norms
 - B) are always less useful than national norms
 - C) are only applicable to school grade children
 - *D) can be especially useful in some situations
68. Psychometric is to edumetric, as _____ is to _____ .
- A) stanine; z score
 - B) item difficulty; item discrimination
 - *C) norm referenced; criterion referenced
 - D) unit weighting; differential weighting
69. We administer to Bill five different tests of depression. We now wish to obtain an overall test score. The most correct way would be to:
- A) simply add Bill's five raw scores into a total
 - *B) change each of Bill's five raw scores into z scores and add these
 - C) multiply each of Bill's raw scores by the length of the test, so longer tests receive greater weight
 - D) compute Bill's mean score, and change that mean score into a T score
70. I have a 20 item vocabulary test. If I use differential weighting to obtain a total score I would
- *A) score each item according to its difficulty
 - B) award 1 point per correct item
 - C) score each item according to its position (e.g., item 1 gets 1 point, item 8 gets 8 points)
 - D) have an ipsative measure
71. Multiple cutoff procedures
- A) allow poor performance to be compensated by good performance
 - *B) do not allow superior performance on one variable to compensate for poor performance on another
 - C) require that all variables be expressed in T scores
 - D) can be applied only to dichotomous decisions
72. A multiple regression
- *A) is a compensatory model
 - B) can only handle four or fewer variables
 - C) does not take into account how variables correlate with each other
 - D) cannot be used to predict a particular outcome
73. The variables that enter a regression equation are those that correlate the _____ with the criterion and the _____ among themselves.
- A) highest; highest
 - *B) highest; lowest
 - C) lowest; lowest
 - D) lowest; highest

74. Discriminant analysis differs from multiple regression in that
- A) the predicted variable is continuous
 - *B) the predicted variable is nominal
 - C) the variables that enter the discriminant analysis are uncorrelated
 - D) we need to know the item discrimination index for each variable

3 Reliability and Validity

1. Reliability can be simply defined as
 - A) accuracy
 - *B) consistency
 - C) truthfulness
 - D) variation
2. Lack of reliability equals
 - *A) error variance
 - B) true variance
 - C) interitem consistency
 - D) degree of homogeneity
3. Test-retest reliability involves _____ group(s) of individuals assessed with _____.
 - A) one; two different tests
 - B) two; two different tests
 - *C) one; one test
 - D) two; one test
4. It is expected that reliability coefficients be at least
 - A) .30
 - B) .50
 - *C) .70
 - D) .90
5. A very short test-retest interval may yield a(n)
 - A) spuriously low coefficient
 - *B) spuriously high coefficient
 - C) unreliable coefficient
 - D) very valid coefficient
6. A long test-retest interval may reflect
 - A) lack of validity in a test
 - B) the degree of interitem consistency
 - C) the standard error of measurement
 - *D) real changes that have taken place
7. Usually, test-retest reliability is assessed
 - *A) over a short period of time
 - B) over a long period of time
 - C) by the degree of interitem consistency
 - D) through the use of alternate forms

8. Several of the problems associated with test-retest reliability reflect the fact that
 - *A) people are different
 - B) males are more stable than females
 - C) individuals are likely to cheat when bored
 - D) contrasted groups are difficult to locate
9. In reporting test-retest reliability we should also report
 - A) the size of the Spearman-Brown coefficient
 - B) the standard error of measurement
 - C) the standard error of differences
 - *D) the length of the interval
10. Item sampling can lower _____ reliability.
 - *A) alternate form
 - B) test-retest
 - C) primary
 - D) interitem consistency
11. The major problem with alternate form reliability:
 - A) motivation on the part of the subject
 - B) lack of interitem consistency
 - *C) developing an alternate form
 - D) degree of bandwidth and fidelity
12. The focus of both _____ and _____ reliability is on internal consistency.
 - A) test-retest; split-half
 - *B) split-half; interitem consistency
 - C) equivalent forms; interitem consistency
 - D) alternate forms; test-retest
13. There are many ways to split a test in two, but in general the _____ method usually works well.
 - *A) odd versus even
 - B) first half versus second half
 - C) easy versus difficult
 - D) homogeneous versus heterogeneous
14. Test A is composed of 200 items, while Test B is composed of 50 items. Both tests have a split-half reliability of .68. Using the Spearman-Brown formula, which test will be more reliable?
 - A) Test A
 - B) Test B
 - *C) both will be equal
 - D) need additional information to tell
15. A test has a reliability of .62. If I increase the length of the test with similar items, I will
 - A) decrease its reliability
 - *B) increase its reliability
 - C) decrease its validity
 - D) increase its validity
16. A basic assumption of the Spearman-Brown formula is that
 - A) all the test items are of comparable difficulty
 - *B) items that are added or eliminated are of equal reliability
 - C) the variance of scores is always zero
 - D) criterion contamination is held in check

17. The Rulon formula can be used to assess
 - *A) split-half reliability
 - B) concurrent validity
 - C) alternate forms reliability
 - D) the degree of specificity
18. In effect, reliability is a function of the _____ of the scores.
 - A) inconsistency
 - B) differential validity
 - *C) variability
 - D) sensitivity
19. From a logical point of view, interitem consistency is closest to:
 - A) test-retest
 - B) Kappa coefficient
 - C) pattern matching
 - *D) split-half
20. As a general rule, the _____ in test scores, the better off we are psychometrically.
 - *A) greater the variability
 - B) lesser the variability
 - C) smaller the individual differences
 - D) larger the standard error of measurement
21. An interitem consistency reliability coefficient of .40 may indicate that the test is
 - A) completely useless
 - B) totally invalid
 - *C) made up of heterogeneous items
 - D) made up of homogeneous items
22. Interitem consistency can be assessed by the _____ formula.
 - A) Spearman-Brown
 - *B) Kuder-Richardson
 - C) Rulon
 - D) generalizability
23. Generalizability theory differs from classical reliability theory in that
 - *A) sources of error are seen as something to be studied
 - B) reliability is assumed to reflect only "true" scores
 - C) sensitivity is seen as more important than specificity
 - D) bandwidth is assumed to be error
24. In assessing scorer or rater reliability, we need to take into account
 - A) whether the raters are all of the same gender
 - *B) agreement above that due to chance
 - C) alternate forms reliability
 - D) pattern matching
25. The _____ is used to assess rater reliability.
 - A) Spearman-Brown formula
 - B) Rulon formula
 - C) coefficient alpha
 - *D) Kappa coefficient
26. The correction for attenuation involves assuming
 - A) perfect validity
 - B) above chance agreement
 - *C) perfect reliability
 - D) a zero standard error of differences

Reliability and Validity

27. The _____ assesses reliability from the individual point of view.
- *A) standard error of measurement
 - B) standard error of estimate
 - C) Kappa coefficient
 - D) coefficient alpha
28. Given a test with standard error of measurement equal to 2; if Rebecca obtains a score of 60, we can say that the probability of her true score being between _____ and _____ is 94%.
- A) 58, 62
 - B) 54, 66
 - *C) 56, 64
 - D) 40, 80
29. The standard error of differences is
- A) equivalent to the Kappa coefficient
 - B) applied only when we use the Spearman-Brown formula
 - *C) basically the sum of two standard errors of measurement
 - D) computed when one of the two tests is not reliable
30. The reliability of difference scores is _____ the sum of the reliability of the two scores.
- A) equal to
 - *B) not equal to
 - C) always smaller than
 - D) always larger than
31. A perfectly reliable test _____ valid.
- *A) may be
 - B) is always
 - C) is never
 - D) none of these
32. _____ validity refers to the question of whether a test adequately covers a particular domain.
- A) Concurrent
 - B) Face
 - *C) Content
 - D) Criterion
33. _____ validity is not considered true validity by some experts.
- *A) Content
 - B) Construct
 - C) Criterion
 - D) Concurrent
34. Smith postulates that happiness is composed of four dimensions, and her test of happiness in fact has four factors. This might be seen as evidence for:
- A) concurrent validity
 - *B) content validity
 - C) high fidelity
 - D) high sensitivity
35. A taxonomy or blueprint can be useful in developing _____ validity.
- *A) content
 - B) concurrent
 - C) primary
 - D) secondary

36. _____ validity involves comparing test scores with some standard.
- A) Construct
 - B) Pattern matching
 - *C) Criterion
 - D) Discriminant
37. The main purpose of _____ is to develop a test as a substitute for a more time consuming or expensive procedure.
- A) test-retest reliability
 - *B) concurrent validity
 - C) coefficient Kappa
 - D) meta analysis
38. All workers at plant X are given an IQ test and the results made available to their supervisors. We now administer a new intelligence test to all the workers, and validate the results against supervisors' ratings. We might be concerned with:
- A) the construct validity of the original test
 - *B) criterion contamination
 - C) cross-validation
 - D) convergent validity
39. _____ is an umbrella term that encompasses information about a particular test as long as it is theory driven.
- *A) Construct validity
 - B) Discriminant validity
 - C) The multitrait-multimethod matrix
 - D) Meta analysis
40. _____ involves correspondence between theory and observed data.
- A) Convergent validity
 - B) Criterion contamination
 - C) Predictive validity
 - *D) Pattern matching
41. Which of these statements is true?
- A) The approaches of content and criterion validity ask different questions, but the results are identical.
 - *B) The approaches of content and criterion validity ask different questions, and the results may be quite different.
 - C) The multitrait-multimethod matrix is to validity as criterion contamination is to reliability.
 - D) Coefficient kappa is used when we wish to estimate the internal consistency of a multiple choice test.
42. Which of these methods does not belong as a way to assess construct validity?
- A) correlation
 - B) group differences
 - C) change over occasions
 - *D) criterion contamination
43. We administer two tests of depression and two tests of ego-strength to a sample. A low or zero correlation between the measures of depression and the measures of ego-strength could be considered evidence for _____ validity.
- *A) discriminant
 - B) convergent
 - C) concurrent
 - D) secondary

Reliability and Validity

- 44. A test without face validity
 - A) cannot be valid
 - *B) can be valid
 - C) also lacks construct validity
 - D) cannot be reliable
- 45. A test predicts well the academic achievement of high school students but not of college students. The test is said to have
 - A) face validity
 - B) poor construct validity
 - *C) differential validity
 - D) inadequate reliability
- 46. We are interested in the validity of the Jones test of mechanical aptitude. We discover that there are 45 published studies that used the test. To come to an overall conclusion about the validity of this test we would use
 - A) tertiary validity
 - B) the multitrait-multimethod matrix
 - C) an expectancy table
 - *D) meta analysis
- 47. I have developed a test of anxiety which contains 560 T-F items. We could characterize this test as having
 - A) homogeneity of items
 - *B) high fidelity
 - C) extremely high internal consistency
 - D) low discriminant validity
- 48. Screening tests designed to assess the presence or absence of signs of mental illness most likely have
 - A) poor primary validity
 - B) high discriminant validity
 - *C) broad bandwidth
 - D) low standard errors of measurement
- 49. All other things being equal, _____ results in a lowered correlation between test scores and criterion.
 - A) greater variability
 - *B) lesser variability
 - C) criterion contamination
 - D) a large standard error
- 50. Tests with low validity
 - *A) may be useful under some circumstances
 - B) cannot have adequate reliability
 - C) should not be used
 - D) can be used only if their SE of estimate is large
- 51. A potential problem with expectancy tables is that
 - A) very large samples need to be studied
 - *B) a specific prediction may be based on few cases
 - C) they do not take into account the validity of the test
 - D) they result in too many false positives

52. We use the SAT to decide whether John should be admitted to Cucamonga University. His score of 650 indicates he can do excellent work, but in fact he flunks out. In decision theory terminology, John is considered
- A) false positive
 - *B) false negative
 - C) false success
 - D) true positive
53. If we wish to decrease the number of false positives in a situation like college admission to a community college, we would
- *A) lower our admission standards
 - B) increase our admission standards
 - C) admit only the top 10% of applicants
 - D) none of these
54. The Yale test of schizophrenia is quite accurate in identifying schizophrenics. The test is said to have high
- A) specificity
 - B) false negatives
 - *C) sensitivity
 - D) false positives
55. The Harvard test of schizophrenia has a very high rate of true negatives; therefore it would
- *A) diagnose as schizophrenics very few normals
 - B) diagnose as schizophrenics most schizophrenics
 - C) diagnose as normal most schizophrenics
 - D) yield a large number of false positives
56. Ideally, we would want a test to have _____ specificity and _____ sensitivity.
- *A) high; high
 - B) high; low
 - C) low; low
 - D) low; high
57. In the example discussed in the text in which a large sample of psychiatric patients were studied for five years to determine whether suicide could be predicted, the results indicated
- A) high sensitivity but very low specificity
 - B) very high predictive value but low sensitivity
 - *C) good sensitivity and specificity, but low predictive value
 - D) very low sensitivity, specificity, and predictive value
58. How can we reduce decision errors?
- A) by using more valid procedures
 - B) by using multiple sources of data
 - C) by using sequential strategies
 - *D) all of these
59. Students who take the Spanish test and obtain a raw score of 108 or above are placed in intermediate Spanish. The raw score of 108 is called a
- A) placement score
 - B) false positive score
 - *C) cutoff score
 - D) selection score
60. The number of individuals we need to select from a pool of applicants:
- *A) selection ratio
 - B) base rate
 - C) sensitivity
 - D) differential validity

Reliability and Validity

61. Despite the newspaper headlines, very few individuals attempt suicide in jail. If I were to screen all inmates with the Rorschach to identify potential suicide attempters, I would commit a lot of errors. This is because of:
- A) selection ratio
 - *B) base rate
 - C) poor specificity
 - D) low primary validity
62. In regards to sample size and correlation coefficient: the _____ the sample, the _____. .
- *A) larger, smaller a coefficient need be to be significant
 - B) smaller, greater the selection ratio
 - C) larger, larger a coefficient need be to be significant
 - D) smaller, higher the base rate
63. In Gough's model, primary validity is basically similar to _____ validity.
- A) construct
 - B) concurrent
 - *C) criterion
 - D) differential
64. Test X is presented by its authors as a measure of self-esteem. The fact that all of the test items are written in a positive manner (e.g., I am a happy person; I like myself) would be considered
- A) primary validity
 - *B) secondary validity
 - C) tertiary validity
 - D) none of these
65. Test X, a measure of self-esteem, is based on the theoretical writings of Carl Rogers.
- A) primary validity
 - *B) secondary validity
 - C) tertiary validity
 - D) none of these
66. Test X is a measure of self-esteem. Women who experience post-partum depression score higher on test X.
- A) primary validity
 - B) secondary validity
 - *C) tertiary validity
 - D) none of these
67. Which of these does not belong?
- A) reviewing the theory and procedures of how a test was developed
 - B) a logical analysis of the item content
 - *C) applying factor analysis to a test
 - D) correlating test scores with important variables
68. The College Prediction Scale is a new test designed to compete with the SAT as to predicting who will do well academically in college. As part of primary validity we would consider:
- A) test-retest reliability
 - B) the theoretical rationale for the test
 - C) whether the test shows gender differences
 - *D) the correlation between test scores and GPA

4 Personality

1. A typical definition of personality would contain the words:
 - *A) variety and unique organization
 - B) extraversion and introversion
 - C) social desirability and uniqueness
 - D) interpersonal style and orientation
2. Mischel (1968) in his textbook argued that one should pay attention to
 - A) the factor structure of personality
 - B) how many basic types of personality exist
 - C) both genetic and developmental aspects
 - *D) the interaction of person and situation
3. Mischel's (1968) argument:
 - A) was totally disregarded by psychologists
 - *B) had been made by others as well
 - C) resulted in the construction of the 16PF
 - D) gave rise to a whole generation of tests
4. The psychodynamic model of personality assumes that
 - *A) early experiences affect later development
 - B) situations are the main source of behavioral differences
 - C) introspection and subjective experiences are most important
 - D) personality develops because of the "self-concept"
5. The situational model assumes that
 - A) if you change the situation, the behavior stays the same
 - B) early experiences affect later development
 - *C) situations are the main source of behavioral differences
 - D) traits are extremely stable
6. Are self-ratings more valid than questionnaires?
 - A) questionnaires are more valid
 - *B) self-ratings are more valid
 - C) both are equally valid
 - D) only for males
7. Self-report measures assume that the respondent
 - *A) is probably the best observer of his/her own behavior
 - B) will fake if given the opportunity
 - C) answers in a highly truthful manner
 - D) will give responses that deviate from the norm

8. To identify the extent to which the respondent presents a biased picture on a personality test, we can use
 - A) reliability scales
 - B) factor analysis
 - C) folk concepts
 - *D) validity scales
9. Projective techniques typically involve
 - A) a set of inkblots
 - *B) an ambiguous set of stimuli
 - C) true-false statements
 - D) multiple choice items
10. The labels attached to a rating scale (such as “above average”) are called:
 - A) rating values
 - B) folk concepts
 - *C) anchor points
 - D) desirability ratings
11. Observing a group of managers in a leaderless group discussion:
 - *A) situational method
 - B) protective technique
 - C) rating scale
 - D) phenomenological assessment
12. Behavioral assessment
 - A) sees the content of self-report as irrelevant
 - *B) does not assume the existence of traits
 - C) is based on the use of projective techniques
 - D) assumes the presence of a basic personality core
13. The trait approach assumes that
 - A) personality traits do not predict behavior accurately
 - B) major changes in personality occur as people get older
 - *C) personality is marked by stability and regularity
 - D) all of these
14. A(n) _____ is a category of individuals all of whom presumably share a combination of _____.
 - *A) type; traits
 - B) body humor; body fluids
 - C) trait; folk concepts
 - D) ipsative scale; responses
15. A phlegmatic individual can be described as:
 - A) depressed
 - B) irritable
 - C) happy
 - *D) apathetic
16. The method of internal consistency, as a way of developing personality tests, focuses on:
 - A) reliability
 - *B) factor analysis
 - C) the use of theory
 - D) folk concepts
17. The _____ was developed using the internal consistency approach.
 - A) California Psychological Inventory
 - B) Myers-Briggs Type Indicator
 - *C) 16 Personality Factors Inventory
 - D) Inventory of Psychosocial Balance

18. From the point of test construction, which of these does NOT belong?
 - A) Myers-Briggs Type Indicator
 - B) Edwards Personal Preference Schedule
 - *C) California Psychological Inventory
 - D) Personality Research Form
19. The method of criterion keying
 - A) always uses contrasted groups
 - B) focuses on theory
 - *C) is marked by validity-in-use
 - D) is also known as the rational approach
20. The fiat method focuses on
 - A) the results of factor analyses
 - B) the empirical correlation with a criterion
 - C) the degree of internal consistency
 - *D) the author's judgment and expertise
21. From a personality point of view, language is important because:
 - A) as the person matures, language becomes more important
 - *B) language encodes important experiences
 - C) personality test items use language
 - D) it is the only way to describe traits
22. Much of the focus of personality is on psychopathology because:
 - A) pathology is easier to assess than normality
 - B) of the impact of the phenomenological approach
 - C) the MMPI has dominated the field
 - *D) many personality theories were developed in the context of clinical patients
23. These tests are partly based on Murray's theory:
 - *A) Edwards Personal Preference Schedule; Personality Research Form
 - B) Personality Research Form; California Psychological Inventory
 - C) Myers-Briggs Type Indicator; Boredom Proneness Scale
 - D) Inventory of Psychosocial Balance; Color Pyramid Test
24. Murray's concept of presses is most akin to:
 - A) Freud's focus on early developmental stages
 - B) Gough's notion of folk concepts
 - C) Cattell's primary traits
 - *D) Mischel's focus on situations
25. The 16 PF is based on the notion that:
 - *A) there are 16 basic personality dimensions scaled through factor analysis
 - B) personality is pretty much established once a person reaches age 16
 - C) there are 16 traits of pathology
 - D) there are 16 basic personality types
26. Which of these is not correct in regards to the 16 PF?
 - A) the 16 PF has been revised a number of times
 - *B) test items typically are scored for several scales
 - C) there are several forms of the test
 - D) the 16 dimensions are said to be independent
27. The 16 PF
 - *A) is basically a self-administered test
 - B) requires a skilled examiner to administer
 - C) yields scores on three vector scales
 - D) is based on the theory of E. Erikson

28. A sten is a standard score
- A) with a mean of 10
 - B) that can range from 0 to 100
 - C) with a mean of 50 and a SD of 10
 - *D) with a mean of 5.5
29. Which of these does NOT apply to the 16 PF:
- A) the test can be hand scored or machine scored
 - *B) the raw scores are changed to T scores
 - C) the items are forced choice items
 - D) good descriptions are given for low and high scorers
30. Computer scoring services for the 16 PF:
- A) are not available at this point
 - B) are available, but not for all the scales
 - C) are not available, except for research purposes
 - *D) are available, some geared for specific purposes
31. In regards to the reliability of the 16 PF:
- A) alternate form reliability is extremely high
 - B) test-retest reliability is not reported
 - *C) internal consistency is on the low side
 - D) alternate form reliability cannot be computed
32. The test manual of the 16 PF gives considerable evidence as to _____ validity.
- *A) factorial
 - B) concurrent
 - C) divergent
 - D) predictive
33. The norms on the 16 PF can be described as:
- A) basically samples of convenience
 - B) local norms only
 - C) based on small, atypical samples like football players
 - *D) extensive and stratified according to census data
34. The scales on Cattell's Sixteen Personality Factor Questionnaire:
- *A) do correlate with each other, some rather substantially
 - B) do correlate with each other, but minimally
 - C) do not correlate with each other since factor analysis was used
 - D) are statistically independent of each other
35. Secondary traits are the result of
- A) the application of secondary validity
 - *B) factor analysis of the primary traits
 - C) gender differences on the original traits
 - D) the use of regression equations
36. A criticism of the 16 PF:
- A) little validity information is available
 - B) it has not been revised since its inception
 - *C) regression equations have not been replicated
 - D) little research on the test is available
37. The Myers-Briggs Type Indicator
- A) was developed by the criterion keying method
 - *B) is based on Jungian theory
 - C) clearly reflects a situational model
 - D) is based on Murray's theory of needs

38. _____ is the process of coming to conclusions.
A) Perception
B) Introversion
*C) Judgment
D) Sensing
39. The two aspects of perception are:
A) thinking and feeling
B) thinking and sensing
C) feeling and intuition
*D) sensing and intuition
40. Jungian types are presumed to be
*A) fairly stable
B) quite unstable
C) a function of the situation
D) biochemically based
41. Both the Cattell 16 PF and the MBTI:
A) used contrasted groups in their development
*B) use forced choice items
C) are based on factor analysis
D) use true-false inventory items
42. The MBTI is applicable to:
A) developmentally disabled children
B) older mature adults only
*C) high school through adult respondents
D) normal school children only
43. A result on the MBTI might look like this:
A) a T score of 68 on introversion
B) a sten of 8 on thinking-feeling
C) higher on judging than intuition
*D) ESFP
44. In regards to the reliability of the MBTI:
*A) adequate and similar to that of other personality inventories
B) clearly marginal at best
C) much higher alternate forms reliability
D) poor test-retest but adequate split-half
45. In regards to the validity of the MBTI:
A) has not been investigated
*B) basically supportive of the construct validity
C) has face validity but little else
D) based primarily on the multitrait-multimethod approach
46. Which of these applies to the MBTI?
A) a good example of the criterion-keying method
B) has three vector scales
*C) basically an ipsative instrument
D) can be scored for both primary and secondary factors
47. Which of these types tend to go together?
A) extraversion and judging
B) thinking and intuition
C) sensation and judging
*D) intuition and perceiving

48. A criticism of the MBTI:
- *A) how well does it mirror Jungian theory
 - B) how stable is its factor structure
 - C) poor test-retest reliability
 - D) inadequate norms
49. Social desirability influenced the development of the
- *A) EPPS
 - B) MBTI
 - C) CPI
 - D) 16 PF
50. The scales of the EPPS reflect:
- A) types from Jungian theory
 - *B) needs from Murray's theory
 - C) folk concepts
 - D) Eriksonian stages of life
51. Both the EPPS and the MBTI
- A) assess Jungian types
 - B) have time limits
 - *C) are ipsative measures
 - D) have poor internal consistency
52. In regards to the EPPS:
- A) there is voluminous evidence supportive of its validity
 - B) the test appears to have low reliability
 - C) norms are not available
 - *D) there is little direct evidence of its validity
53. The initial normative sample on the EPPS:
- *A) male and female college students
 - B) community adults chosen on the basis of U.S. census
 - C) various samples of convenience, like police officers
 - D) no normative samples were used
54. On the EPPS, a consistency score can be determined by
- A) a set of questions that focus on consistency
 - B) a test-retest procedure
 - C) assessing the split-half reliability
 - *D) comparing the responses to 15 items that are repeated
55. On the EPPS, the index of profile stability
- A) is essentially a test-retest reliability coefficient
 - *B) is essentially a split-half correlation coefficient
 - C) can be read directly from a table
 - D) is based on 15 items that are repeated
56. If one were to use the consistency score on the EPPS to determine invalid profiles, approximately _____ % of the normative protocols would be deemed invalid.
- A) 5
 - *B) 25
 - C) 50
 - D) 75
57. Items used in the EPPS:
- A) all reflect Jungian theory
 - B) are primarily true-false
 - C) were retained because of factorial validity
 - *D) were equated in social desirability

58. Both the EPPS and the Personality Research Form:
 - *A) are based on the need theory of Murray
 - B) are ipsative instruments
 - C) use forced choice items
 - D) require a skilled examiner to administer
59. Which one of these is NOT a basic principle that guided the construction of the Personality Research Form?
 - *A) items were retained that conformed to the author's theory
 - B) each of the traits was explicitly and theoretically defined
 - C) there was a large item pool
 - D) social desirability was controlled for
60. The scales on the Personality Research Form
 - A) have generally low test-retest reliability
 - *B) have high internal consistency
 - C) assess such traits as intuition and extraversion
 - D) are highly susceptible to social desirability
61. Alternate form reliability could be computed on all of these except the:
 - A) MBTI
 - *B) CPI
 - C) 16 PF
 - D) PRF
62. If we were looking for a test that presumably measures the same variables as the EPPS, we would select the
 - A) MBTI
 - B) 16 PF
 - *C) PRF
 - D) CPI
63. If we were looking for a scale of social desirability we might select the
 - A) CPI
 - B) Color Pyramid Test
 - C) EPPS
 - *D) PRF
64. The infrequency scale of the PRF is composed of items
 - *A) for which there is a clear typical answer
 - B) for which there is no clear typical answer
 - C) that are repeated throughout the test
 - D) that are rarely answered as true by anyone
65. Most of the validity data presented by the author of the PRF reflects _____ validity.
 - *A) convergent
 - B) discriminant
 - C) factorial
 - D) predictive
66. The PRF is characterized as:
 - A) having excellent validity but poor reliability
 - *B) highly sophisticated in its development
 - C) providing substantial norms based on national samples
 - D) having a rather unique theoretical rationale

67. A major criticism of the PRF:
- A) too long
 - B) no norms for college students
 - *C) lack of validity studies
 - D) does not consider social desirability
68. What should be measured in a personality test? For Gough the answer is:
- *A) folk concepts
 - B) primary traits
 - C) Jungian types
 - D) needs and presses
69. "Folk concepts" can best be described as:
- A) secondary traits resulting from factor analysis
 - B) internally consistent and independent dimensions
 - C) traits that are specific to a particular culture
 - *D) important dimensions of personality embedded in language
70. CPI scales have two major purposes, one of which is:
- *A) to predict what people will say and do in specific contexts
 - B) to assess the basic needs that people have
 - C) to identify how intuitive a person is
 - D) to determine how valid self-ratings are
71. In regards to the statement that personality scales should be factorially pure, Gough would reply:
- A) factor analysis is the way to develop scales
 - *B) independence is a preference and not a law of nature
 - C) an open system does not need factor analysis
 - D) that's exactly what the CPI has
72. The five "axioms" that underlie the understanding of the CPI
- A) reflect a strong factor analytic bias
 - B) are embedded in Jungian theory
 - *C) can be asked of any personality inventory
 - D) are the same dimensions of the "Big Five" model
73. The CPI scales
- A) were developed strictly from an empirical point of view
 - *B) are in many instances the result of a theoretical framework
 - C) are strictly atheoretical
 - D) parallel the scales of the MMPI
74. The basic 20 scales of the CPI fall into four groups. Which one does NOT belong?
- A) interpersonal style
 - *B) introversion-extraversion
 - C) normative values and orientation
 - D) cognitive-intellectual functioning
75. Folk concepts are presumed to be
- A) genetically determined
 - B) a reflection of our Anglo-Saxon heritage
 - *C) found in all cultures
 - D) secondary traits
76. On the CPI we would expect that test items
- *A) are often scored on more than one scale
 - B) would be scored for only one scale
 - C) would be scored for all scales
 - D) do not correlate highly with each other

77. On the CPI, alphas are
 - A) skeptics and innovators
 - B) introverted and intuitive
 - *C) leaders and doers
 - D) visionary and maladapted
78. Which of these is NOT a vector scale of the CPI?
 - A) norm accepting - norm questioning
 - B) self-realization and fulfillment
 - C) introversion-extraversion
 - *D) need Achievement
79. The CPI scales are listed on its profile
 - A) in alphabetical order
 - *B) in order of their psychological relationship to each other
 - C) most reliable on the left, least reliable on the right
 - D) according to the subject's T scores
80. The reliability of the CPI scales
 - *A) is acceptable for some, and marginal for others
 - B) is exceptionally high for all the scales
 - C) consists of test-retest reliability coefficients only
 - D) is marginal at best
81. Gough's nationwide normative study of high school students found that
 - A) alphas tended to major in art and music
 - B) deltas tended to major in business and medicine
 - *C) deltas had the lowest rate of college entrance
 - D) alphas had the lowest rate of college entrance
82. The CPI has been applied to
 - *A) a wide range of topics
 - B) the study of academic achievement only
 - C) high school students only
 - D) the study of pathology only
83. The 20 folk concept scales of the CPI
 - A) do not intercorrelate
 - *B) correlate positively with each other
 - C) in general correlate negatively with each other
 - D) represent the 20 needs postulated by Murray
84. A factor analysis of the CPI would most likely yield _____ factor(s).
 - A) one
 - B) two
 - *C) four
 - D) twenty
85. According to Erikson, the challenge of identity occurs in
 - A) early infancy
 - B) early childhood
 - C) middle childhood
 - *D) adolescence
86. According to Erikson, the challenge to be met in early infancy is:
 - *A) trust vs. mistrust
 - B) autonomy vs. shame
 - C) initiative vs. guilt
 - D) intimacy vs. isolation

87. The items in the Inventory of Psychosocial Balance differ from most other personality inventories in that they use a(n) _____ format.
- A) multiple choice
 - *B) Likert
 - C) matching
 - D) odd-man-out
88. The initial item pool of the Inventory of Psychosocial Balance was initially reduced by using:
- A) factor analysis
 - B) criterion-keying
 - *C) psychologists' judgments
 - D) regression equations
89. The alpha coefficients for the Inventory of Psychosocial Balance are:
- A) below acceptable limits
 - B) extremely high
 - C) not computable
 - *D) acceptable but low
90. In the validity study of the IPB where 18 highest scoring and 18 lowest scoring female college students were interviewed:
- A) low scorers were seen as psychologically healthy
 - *B) high scorers were seen as psychologically healthy
 - C) all subjects were seen as socially mature
 - D) all subjects were seen as having major personality problems
91. Correlations between IPB scales and variables like social desirability and intelligence suggest that
- A) the convergent validity of the IPB is excellent
 - B) the construct validity of the IPB is inadequate
 - *C) the discriminant validity of the IPB is good
 - D) the concurrent validity of the IPB is quite good
92. At present time, the norms on the IPB consist of:
- *A) small samples from several studies
 - B) large samples chosen with census data
 - C) both random and representative groups
 - D) various groups from children to adults
93. In a study of elderly persons with the IPB, it was found that
- A) there were no significant gender differences
 - B) the IPB was not valid with elderly subjects
 - C) most subjects had not resolved their identity crises
 - *D) for men, higher effective functioning was related to trust and industry
94. The Self-Consciousness Scale was developed
- *A) by writing items that matched a theoretical framework
 - B) through the use of contrasted groups
 - C) by retaining items that correlated with an outside criterion
 - D) by the use of criterion-keying
95. The Self-Consciousness Scale yields
- A) a total score only
 - *B) scores on 3 factors and a total score
 - C) scores on seven aspects of self-consciousness
 - D) scores on five factors

96. Scoring of the Self-Consciousness Scale
 - A) is a complicated matter needing computer use
 - B) requires a well trained examiner
 - *C) reflects unitary weights
 - D) reflects differential weighting
97. Criticisms of the Self-Consciousness Scale center mostly on its
 - A) reliability
 - *B) validity
 - C) norms
 - D) factor structure
98. Theoretically, we would expect scores on a boredom proneness scale to be correlated with measures of
 - *A) depression and loneliness
 - B) industry and intimacy
 - C) extraversion and introversion
 - D) need aggression
99. The Boredom Proneness Scale was included in the text because it illustrates
 - A) the factor analytic approach
 - *B) the basic steps to be followed in developing a scale
 - C) the internal consistency approach
 - D) how excellent norms can be gathered with little effort
100. The Color Pyramid Test differs from the other personality measures covered in that:
 - A) it was developed in the U.S. and validated in Germany
 - B) the scoring is strictly impressionistic
 - *C) it is not a paper-and-pencil test
 - D) it has a strict time limit
101. How many personality dimensions are there? A popular answer is:
 - *A) 5
 - B) 16
 - C) 20
 - D) over 100
102. The letters NEO stand for:
 - A) Normality, Exceptionality, Orderliness
 - B) the three authors (Norman, Everett, & Oserinsky)
 - C) New England Observational (scale)
 - *D) Neuroticism, Extraversion, Openness to experience
103. The forms of the NEO-PI-R differ from the forms for most other personality inventories in that:
 - *A) one form is a self-report form, the other an observer report
 - B) they are very short yet highly reliable
 - C) the forms contain different scales
 - D) there are separate forms for males and for females
104. Most commercially published personality inventories
 - A) incorporate the answers into an expendable booklet
 - *B) have reusable booklets and machine scored answer sheets
 - C) have reusable answer sheets
 - D) have scale items presented in sequence within the inventory

105. The NEO-PI-R yields
- *A) 30 facet scores and 5 domain scores
 - B) 30 domain scores and 5 facet scores
 - C) 5 basic scores and a total
 - D) 16 scores based on factor analysis
106. Reviews of the NEO-PI commend it highly because
- A) it has extensive norms
 - B) its five scales are highly independent
 - C) the facet scales are highly independent
 - *D) it has good discriminant and convergent validity
107. We would find scales like “Threctia-Parmia” on the
- A) CPI
 - B) MBTI
 - *C) 16 PF
 - D) NEO-PI
108. If we wanted to measure need-endurance we might consider the
- A) CPI
 - *B) EPPS
 - C) Color Pyramid Test
 - D) MBTI
109. On the CPI a person labelled as a “beta type” would score toward the
- *A) rule accepting and introverted poles
 - B) rule questioning and introverted poles
 - C) rule accepting and extraverted poles
 - D) rule questioning and extraverted poles

5 Cognition

1. The changes that are occurring in the assessment of cognitive abilities are consonant with the primacy of _____ validity.
A) concurrent
*B) construct
C) primary
D) factor
2. A typical achievement test assesses what a person
*A) has learned over a specific time period
B) has learned over their lifetime
C) is potentially able to achieve
D) can learn in a reasonable amount of time
3. The progression from a global conception of intelligence to one of various components, was in part due to
*A) factor analysis
B) Spearman's theory
C) the development of point scales
D) the concept of scatter
4. Geographic metaphors of intelligence
A) define intelligence in terms of brain function
B) look at the philosophical underpinnings
*C) focus on the structure rather than the process of intelligence
D) emphasize the context of culture
5. Most tests of intelligence have their origins in the _____ metaphor of intelligence.
A) biological
B) anthropological
C) sociological
*D) geographic
6. Computational metaphors of intelligence
A) focus on mathematical abilities
B) study patients that have brain lesions
*C) use the computer as a model of the brain
D) look at the philosophical underpinnings
7. The theory of Jean Piaget would fall under which metaphor of intelligence?
A) anthropological
*B) epistemological
C) sociological
D) computational

8. In Piaget's theory, the sensorimotor period is followed by the _____ period.
 - *A) preoperational
 - B) concrete
 - C) formal operational
 - D) epistemological
9. These theories often take a strong negative view of intelligence tests:
 - A) sociological
 - *B) anthropological
 - C) epistemological
 - D) biological
10. "What happens inside the person first happens between people."
 - A) biological metaphor
 - B) epistemological metaphor
 - *C) sociological metaphor
 - D) geographic metaphor
11. Guilford's model of the structure of intellect
 - A) would fall under the biological metaphor category
 - B) is a theoretical model with no practical tests
 - C) postulates multiple intelligences
 - *D) postulates processes, contents, and products
12. Gardner's theory of intelligence
 - *A) postulates seven separate intelligences
 - B) is very much like a factor analytic approach
 - C) suggests that abstract thinking begins at about age 11
 - D) views intelligence in the context of culture
13. In Guilford's model, memory and divergent thinking are categories of
 - A) contents
 - *B) processes
 - C) products
 - D) intelligence
14. A generic criticism of tests like the Stanford-Binet is that they
 - *A) were practical measures not tied to theory
 - B) show very little validity
 - C) are applicable only in the U.S.
 - D) stem from biological theories of intelligence
15. Spearman's two-factor theory sees intelligence as
 - A) composed of multiple abilities
 - *B) a global, unitary ability
 - C) composed of 120 unique combinations
 - D) a set of primary abilities
16. In contrast to the U.S., the more popular approach in England is that of:
 - A) intelligence as made up of multiple factors
 - B) hierarchical theories of intelligence
 - C) the structure of intellect model
 - *D) intelligence as a global, unitary ability
17. How many multiple factors of intelligence are there?
 - A) 120
 - B) 12
 - C) 1
 - *D) no generally agreed number

18. Correlations between general intelligence and job proficiency
 - *A) could be as high as .75
 - B) is typically near zero
 - C) has not been studied to any degree
 - D) cannot be computed because of restriction of range
19. In general, intelligence test scores and academic achievement correlate about:
 - A) zero
 - *B) .50
 - C) .20
 - D) .80
20. The relationship between intelligence test scores and academic achievement is:
 - A) higher in college than primary grades
 - B) equal in magnitude throughout the educational range
 - *C) higher in primary grades than college
 - D) highest in those who go on to graduate school
21. Tacit knowledge refers to knowledge
 - *A) that is practical but not directly taught
 - B) that is directly taught by a mentor
 - C) that the person does not know he/she has
 - D) that the person has not yet mastered
22. Studies have shown that experts versus novices differ in
 - *A) the amount and organization of their tacit knowledge
 - B) their underlying cognitive abilities
 - C) how they solve problems
 - D) how they perform on traditional intelligence tests
23. To the criticism that intelligence test items do not directly measure a person's ability to learn, we can answer:
 - A) that just shows how inadequate intelligence tests are
 - B) tests like the Wechsler series do use such items
 - *C) such correspondence is not necessary
 - D) it's because theories of intelligence do not consider this issue
24. An age scale
 - A) is well illustrated by the Wechsler series
 - *B) uses items placed in age equivalent categories
 - C) shows a mean increase for older children
 - D) assumes a biological basis to intelligence
25. Intelligence Quotient is defined as:
 - A) chronological age divided by mental age, X 100
 - B) the sum of chronological age and mental age, X 100
 - C) the discrepancy between mental age and chronological age, X 100
 - *D) mental age divided by chronological age, X 100
26. A deviation IQ
 - *A) is based on transforming a raw score to a mean of 100 and $SD = 16$
 - B) is identical to the ratio IQ
 - C) equals mental age divided by chronological age
 - D) is applicable only to retarded children
27. To ask whether there are gender differences in intelligence may not be meaningful because:
 - A) the results would be confounded by such aspects as socioeconomic level
 - *B) most tests of intelligence do not use items that show gender differences
 - C) deviation IQs are the same for both genders
 - D) contrasted groups that differ only in gender are not available

28. Tests like the Stanford-Binet and the Wechsler scales
- A) yield only ratio IQs rather than deviation IQs
 - *B) allow for observation of many intellectual processes
 - C) are excellent for prediction but not for understanding
 - D) have substantial face validity but little construct validity
29. Sensitivity represents the proportion of
- A) true positives
 - *B) low performers on the criterion that are accurately identified
 - C) high performers on the criterion that are accurately identified
 - D) true positives to true negatives
30. The Binet-Simon scales were developed as a way to identify:
- A) gifted children
 - B) suicide prone adolescents
 - *C) retarded children
 - D) incompetent adults
31. The 1905 Binet-Simon Scale
- A) yielded a deviation IQ
 - B) required a skilled examiner
 - C) consisted of 30 multiple-choice items
 - *D) was more like a structured interview
32. In comparison to the 1905 Binet-Simon, the 1908 scale allowed the child's score to be expressed as a(n):
- *A) mental age
 - B) IQ
 - C) deviation IQ
 - D) verbal and performance IQ
33. As compared to the two previous versions, the 1911 Binet-Simon scale
- A) had a substantially larger normative sample
 - B) introduced the concept of deviation IQ
 - *C) had only some minor changes
 - D) consisted of two parallel forms
34. The Stanford-Binet differed from the Binet-Simon scales;
- A) in language only
 - *B) in several aspects, including the nature of the examiner
 - C) in being shorter and more streamlined
 - D) in focusing on secondary validity
35. The 1937 Stanford-Binet was criticized because the standardization sample
- *A) was not truly representative
 - B) was quite small
 - C) consisted of California children only
 - D) consisted of gifted children only
36. The 1960 Stanford-Binet
- A) had two separate forms, L & M
 - *B) collapsed two parallel forms into one
 - C) used the ratio IQ
 - D) was applicable only to 12 year olds and younger
37. Increases in intelligence on the 1972 Stanford-Binet were apparently due to
- *A) cultural changes and the impact of television
 - B) lowered alternate form reliability
 - C) inadequate standardization samples
 - D) the inclusion of too many easy items

38. In comparison to earlier versions, the 1986 Stanford-Binet
- A) is essentially identical to earlier versions
 - *B) might as well be considered a new test
 - C) introduced the concept of deviation IQ
 - D) was criticized for being atheoretical
39. The theory that subsumes the 1986 Stanford-Binet
- A) sees intelligence as a global ability
 - B) sees intelligence as composed of five abilities
 - *C) is a hierarchical theory
 - D) does not consider short-term memory
40. Fluid-analytic abilities:
- A) are highly verbal, such as verbal reasoning
 - B) are basically academic skills
 - C) are not considered part of intelligence
 - *D) are non-verbal and involve spatial thinking
41. The 15 subtests of the 1986 Stanford-Binet cover these areas, except for:
- A) verbal reasoning
 - *B) long-term memory
 - C) abstract/visual reasoning
 - D) quantitative reasoning
42. The 1986 Stanford-Binet yields _____ scores.
- *A) 20
 - B) 15
 - C) 4
 - D) 1
43. Administration of the 1986 Stanford-Binet
- A) is a simple matter and requires no special training
 - B) can be learned in a 1 hour training session
 - *C) requires a highly trained examiner
 - D) requires a good knowledge of statistics
44. The 1986 Stanford-Binet is an “adaptive” test. This means that
- A) the test can be given to handicapped children
 - B) the examiner needs to be very flexible
 - C) testing is discontinued when the basal level is reached
 - *D) not all items are administered to all subjects
45. The 15 subtests of the 1986 Stanford-Binet are administered
- A) with the Vocabulary test last
 - *B) in a specific sequence
 - C) with short-term memory test first
 - D) as the examiner deems appropriate
46. The basal level on the 1986 Stanford-Binet
- A) is exactly the same as on other intelligence tests
 - *B) is defined as passing 4 consecutive items
 - C) is defined as 4 consecutive failures
 - D) applies only to the Vocabulary subtest
47. The basal and ceiling levels are used
- A) in order to be able to calculate a deviation IQ
 - B) so as to determine the range of ability
 - *C) to minimize discouragement and maximize information obtained
 - D) because test-retest reliability requires this

Cognition

48. Standard age scores on the Stanford-Binet have a mean of _____ and a SD of _____.
A) 100; 16
B) 50; 10
C) 15; 3
*D) 50; 8
49. The “test composite” score on the Stanford-Binet
*A) is really a deviation IQ
B) is the difference between basal and ceiling level
C) has a mean of 50 and SD of 8
D) has a mean of 500 and SD of 100
50. In regards to the reliability of the Stanford-Binet IV:
A) much of it is based on the Spearman-Brown formula
B) the test seems to have marginal reliability
C) inter-rater reliability has been thoroughly investigated
*D) most is of the internal consistency type
51. In regards to the reliability of the Stanford-Binet IV:
A) the test composite score is reliable but the area scores are not
B) test-retest reliability has not been assessed
*C) most subtests as well as area scores are quite reliable
D) the area scores are reliable but the test composite score is not
52. The validity of the Stanford-Binet IV can best be understood from the aspect of _____ validity:
A) factorial
*B) construct
C) primary
D) content
53. The results of factor analyses of the Stanford-Binet IV indicate that
*A) there is a general factor
B) the factor structure does not vary by age
C) there are 15 basic factors that match the 15 subtests
D) there are 4 primary factors that match the 4 major areas
54. The results of factor analyses of the Stanford-Binet IV
A) do not support the concept of crystallized abilities
*B) do not fully support its theoretical model
C) fully support Binet’s theory of faculties
D) have been obtained only in samples of adults
55. Correlations between total Stanford-Binet scores and similar scores on other tests are typically in the _____ range.
A) .20 to .30
B) .40 to .50
C) .60 to .70
*D) .80 to .90
56. Criticism of the Stanford-Binet IV:
A) items too heavily reflect school experience
B) only few intellectual abilities are tested
*C) standardization sample is not fully representative
D) has no theoretical framework
57. The Wechsler-Bellevue was developed
A) to replace the WAIS
*B) as an adult test of intelligence
C) to directly compete with the Stanford-Binet IV
D) as a test of intelligence for children

58. The initial selection of items for the WAIS was based on:
- A) Spearman's two factor theory
 - B) then current hierarchical theories
 - *C) Wechsler's theory of intelligence
 - D) the structure of intellect model
59. The WAIS-R subtests are divided into:
- *A) Verbal and Performance
 - B) Crystallized and Fluid
 - C) short-term and long-term
 - D) Vocabulary and Quantitative Reasoning
60. On the WAIS raw scores on each subtest are changed into standard scores with a mean of _____ and a SD of _____.
- A) 15; 2
 - B) 50, 8
 - C) 100; 16
 - *D) 10; 3
61. The WAIS-R
- *A) requires a trained examiner to administer
 - B) is easy to administer but difficult to score
 - C) can be administered and scored by computer
 - D) is basically a "quickie" screening instrument
62. On the WAIS-R, IQs of 70 to 79 are labelled as:
- A) superior
 - B) high average
 - *C) borderline
 - D) mentally retarded
63. The Digit Span subtest of the WAIS-R measures
- *A) immediate memory and the disruptive effects of anxiety
 - B) common sense and practical judgment
 - C) the ability to plan
 - D) the ability to perceive part-whole relationships
64. Alertness to details is assessed by the WAIS-R subtest of:
- A) digit symbol
 - *B) picture completion
 - C) information
 - D) arithmetic
65. Which subtest does NOT belong?
- A) Arithmetic
 - B) Vocabulary
 - C) Comprehension
 - *D) Block Design
66. Which subtest does NOT belong?
- *A) Picture completion
 - B) Picture arrangement
 - C) Block Design
 - D) Object assembly
67. In general, the reliability of the WAIS-R
- *A) is excellent
 - B) has not been studied extensively
 - C) is marginal at best
 - D) is inadequate for the individual subtests

Cognition

68. Wechsler argued that the WAIS
- A) is valid because it is empirical
 - *B) has content and construct validity
 - C) is valid because it correlates with GPA
 - D) has face validity and hence primary validity
69. We would expect a typical correlation of about _____ between WAIS-R scores and GPA in college:
- A) .20
 - *B) .40
 - C) .60
 - D) .90
70. The normative sample for the WAIS-R consisted of:
- A) children and adults representing ages 5 to 85
 - B) both normal and psychiatrically impaired adults
 - *C) basically normal adults
 - D) samples of convenience
71. How stable is intelligence over time? According to studies with the WAIS:
- A) generally unstable after age 30
 - B) generally stable but only between ages 20 and 50
 - *C) highly stable for short time periods, and quite stable for longer periods
 - D) stability cannot be separated from validity
72. The Deterioration Quotient
- *A) is based on WAIS subtests where performance changes with age
 - B) is a special score obtained on the K-ABC
 - C) is fully supported by the research literature
 - D) is essentially a deviation IQ
73. Pattern analysis on the Wechsler scales:
- A) is based on the Arithmetic and Comprehension subtests as diagnostic indicators
 - *B) may reflect diagnostic status, low reliability, or other aspects
 - C) is related to cerebral impairment
 - D) is fully supported by the research literature
74. Verbal and Performance IQs on the Wechsler tests correlate about:
- A) .10
 - B) .30
 - *C) .80
 - D) .95
75. Some factor analytic studies suggest one general factor on the WAIS typically called:
- A) short-term memory
 - B) performance
 - *C) general reasoning
 - D) arithmetic reasoning
76. In regards to the factor structure of the Wechsler tests, the literature
- *A) gives equivocal results
 - B) almost unanimously supports one factor
 - C) suggests at least five separate factors
 - D) fully supports the distinction between verbal and performance

77. The subtest reduction method results in an abbreviated or short form of the test, which has
- A) an equal number of subtests as the original
 - B) as many items as the original
 - *C) fewer subtests than the original
 - D) additional items that are more reliable
78. The literature on the Wechsler short forms suggests that:
- *A) the abbreviated forms are useful as screening instruments
 - B) the abbreviated forms are valid IQ measures, but their reliability is poor
 - C) subtest-reduction forms are more valid than item-reduction forms
 - D) in some circumstances the abbreviated forms are more valid than the original
79. The criterion in determining the validity of a short form is
- A) its factor structure
 - *B) its correlation with the full scale score
 - C) its degree of reliability as compared with the original
 - D) based on contrasted groups
80. We would expect a short form of the WAIS to correlate with the Full Scale IQ at about:
- A) .10 to .30
 - B) .30 to .50
 - C) .50 to .60
 - *D) .80 to .90
81. Vocabulary subtest scores on the WAIS-R typically correlate in the _____ with Full Scale IQ.
- A) .30s
 - B) .60s
 - C) .80s
 - *D) .90s
82. Since short forms on the Wechsler tests correlate substantially with the original test, why not use short forms? Short forms _____.
- *A) do not provide as extensive observational information
 - B) have very low reliability because of their shortness
 - C) do not allow the calculation of T scores
 - D) should in fact be used whenever possible
83. In regards to examiner error on the Wechsler scales:
- A) does not occur because the manual is explicit and well written
 - B) can occur but has not been studied
 - *C) occurs because of several reasons, including carelessness
 - D) happens only when there is inadequate training
84. A major criticism of the WAIS-R:
- A) highly biased against minority members
 - *B) not in line with current views of intelligence
 - C) has not been revised since its inception
 - D) has poor primary validity
85. The WISC
- A) contains far fewer subtests than the WAIS
 - *B) is basically a downward extension of the Wechsler-Bellevue
 - C) was modeled on the Stanford-Binet
 - D) has a different theoretical framework than the WAIS

Cognition

86. A national survey of school psychologists indicated that:
- *A) the WISC-R was judged the most valid measure of intelligence
 - B) the K-ABC was preferred over the WISC-R
 - C) school psychologists preferred the WAIS over the WISC
 - D) there were regional differences as to preference
87. The subtests on the WISC-R:
- A) are highly similar to those of the WAIS
 - B) are divided into Verbal and Performance subtests
 - C) include Vocabulary and Information subtests
 - *D) all of these
88. In regards to the administration of the WISC-R:
- A) it is relatively easy and requires little training
 - *B) the starting point on some subtests depends on the child's age
 - C) involves administering all items to every child
 - D) the sequence can be altered to fit the situation
89. On a test like the WISC-R we would generally expect the
- A) split-half reliability to be low
 - *B) reliability of the individual subtests to be lower than that of the total scores
 - C) reliability of the short forms to be substantially lower than that of the original
 - D) test-retest reliability to increase as the time interval gets longer
90. The SE of measurement for the Full Scale IQ on the WISC-R is about
- *A) 3
 - B) 5
 - C) 10
 - D) 15
91. WISC-R scores typically correlate in the _____ range with measures of school achievement such as grades.
- A) 0 to .20
 - *B) .50s-.60s
 - C) .60s-.80s
 - D) above .80
92. Of the various types of validity, fewer studies have looked at the _____ validity of the WISC-R.
- A) concurrent
 - B) construct
 - *C) predictive
 - D) criterion
93. In general, factor analytic studies of the WISC-R
- *A) support the verbal and performance division
 - B) point to "freedom from distractibility" as the major factor
 - C) show a distinct factor structure from the WAIS
 - D) show more factors than subtests
94. Abbreviated scales on the WISC
- A) seem to work better than the original
 - *B) are useful as screening devices
 - C) have not been developed as they have with the WAIS
 - D) correlate rather poorly with the full scale

95. The WPPSI covers the ages of:
A) birth to 4 years
*B) 4 years to 6½ years
C) 4 years to 10 years
D) 5 years to 8 years
96. On the Wechsler tests, raw scores on each subtest are changed to normalized standard scores that have a mean of _____ and SD of _____.
A) 50; 10
B) 100; 15
*C) 10; 3
D) 8; 5
97. Full Scale IQ scores on the WPPSI typically correlate in the .50 to .70 range with scores on other tests. Such correlations should however be _____ because the samples tested are typically _____.
A) lower; homogeneous
*B) higher; homogeneous
C) higher; heterogeneous
D) lower; heterogeneous
98. The WPPSI and the Stanford-Binet correlate so highly with each other; we can therefore conclude that:
A) IQs from one test are interchangeable with the other
B) subtests from one can be used with the other
C) the vocabulary subtest on both is identical
*D) none of these
99. The study that monozygotic twins are closer in intelligence than are dizygotic twins, can be seen as _____ of the WPPSI.
*A) supportive of the construct validity
B) not supportive of the primary validity
C) incongruent with the factor structure
D) supportive of the internal consistency
100. We would expect the correlation between Verbal and Performance IQs to be:
A) lower on the WPPSI than on the WAIS
*B) higher on the WPPSI than on the WAIS
C) basically identical on the WPPSI as on the WAIS
D) to be of the same magnitude on the WPPSI as on the K-ABC
101. The primary purpose of the WPPSI-R is to:
A) assess normality of intellectual functioning
B) compare fluid and crystallized intelligence
*C) diagnose exceptionality
D) measure sequential and simultaneous processing
102. A major criticism of the WPPSI-R:
A) little reliability evidence is available
*B) works well but does not advance our understanding of intelligence
C) too new to have been studied thoroughly
D) does not have construct validity
103. In The United States, the British Ability Scales became the
A) American Ability Scales
B) K-ABC
C) Kohlberg BAS
*D) Differential Ability Scales

Cognition

104. The British Ability Scales is unusual because
- *A) different subtests have different theories
 - B) it does not require a trained examiner
 - C) its reliability is poor, but it has high validity
 - D) it covers the entire age range from “cradle to grave”
105. One aim of the Differential Ability Scales is to:
- *A) produce homogeneous subtests
 - B) measure sequential and simultaneous processing
 - C) yield an overall global composite score
 - D) develop a test that could be administered in pantomime
106. In regards to the Differential Ability Scale, _____ have high loadings on g.
- A) diagnostic subtests
 - B) only information processing subtests
 - *C) core subtests
 - D) all subtests
107. Specificity in the DAS can be defined as:
- A) the amount of variability for the entire test
 - *B) the unique assessment contribution of a subtest
 - C) the abilities measured by a specific subtest
 - D) how subtests correlate with g
108. _____ processing involve stimuli that are primarily _____ .
- A) Sequential; performance
 - B) Fluid; mathematical
 - C) Simultaneous; verbal
 - *D) Simultaneous; spatial
109. The K-ABC was designed to specifically assess the dimensions of:
- A) fluid and crystallized intelligence
 - *B) sequential and simultaneous processing
 - C) primary and secondary processes
 - D) verbal and nonverbal intelligence
110. The Achievement scale of the K-ABC
- A) is part of the Mental Processing Composite scale
 - B) assesses intelligence of children with speech disorders
 - *C) assesses what a child has learned in school
 - D) determines how motivated a child is
111. The reliability of the K-ABC:
- A) does not seem adequate
 - B) has not been studied
 - *C) increases with increasing age
 - D) excellent alternate forms reliability
112. The validity of the K-ABC:
- A) only normal samples have been studied
 - *B) seems quite substantial
 - C) is excellent for retarded children only
 - D) is limited to concurrent studies
113. A new program to identify gifted children is to be started in your local school, with a special effort to identify minority gifted children. You would recommend use of the:
- *A) K-ABC
 - B) WPPSI-R
 - C) Stanford-Binet IV
 - D) Differential Ability Scales

114. The SOI-LA differs from other tests like the Wechsler series and the Stanford-Binet in that:
- *A) it can be administered in a group format
 - B) it lacks a theoretical rationale
 - C) all of its subtests assess sequential processing
 - D) all of its items reflect divergent thinking
115. One of the major criticisms of the SOI-LA:
- A) its high reliabilities yield small SE of measurement
 - B) the factor structure does not match Guilford's theory
 - *C) its low reliabilities yield large SE of measurement
 - D) pays little attention to construct validity
116. The SCAT III is designed to assess
- A) academic learning in the high school years
 - B) vocabulary, especially "street" vocabulary
 - C) primary and secondary thinking in a Freudian framework
 - *D) the accumulation of learning throughout a person's life
117. Unlike the Stanford-Binet and the Wechsler series, the SCAT III is:
- A) a measure of intellectual competence
 - *B) a group test
 - C) well normed
 - D) free of criticism
118. A major criticism of the SCAT III:
- A) requires a trained and skilled examiner
 - B) has poor reliability
 - C) cannot be used with elementary school children
 - *D) lack of information about validity
119. The Otis-Lennon or OLSAT is based on:
- *A) hierarchical theory of intelligence
 - B) Guilford's structure of intellect model
 - C) Cattell's theory of primary factors
 - D) Spearman's two factor theory
120. In regards to the reliability of the Otis-Lennon:
- A) has low internal consistency
 - B) has poor test-retest reliability
 - *C) has high internal consistency
 - D) has not been investigated
121. In regards to the validity of the OLSAT:
- A) has poor content validity
 - B) has inadequate primary validity
 - *C) construct validity is largely absent
 - D) has excellent tertiary validity
122. The Slosson Intelligence Test
- *A) is a brief screening instrument
 - B) requires a well trained examiner
 - C) has a scoring procedure comparable to the WAIS
 - D) takes approximately 60 minutes to administer
123. The Speed of Thinking Test measures
- *A) cognitive speed
 - B) verbal intelligence
 - C) short-term memory
 - D) part-whole relationships

6 Attitudes, Values, and Interests

1. Traditionally, attitudes are seen as having three components:
 - A) primary, secondary, tertiary
 - *B) emotional, behavioral, and cognitive
 - C) attitudes, values, and interests
 - D) direct observation, self-report, and behavior
2. Attitudes are central because they
 - A) are more important than values
 - B) can be measured directly
 - *C) are presumed to reflect behavior
 - D) are highly reliable over time
3. The most obvious way to learn a person's attitudes is to:
 - *A) ask the person
 - B) observe their behavior
 - C) use an attitude scale
 - D) use a sociometric procedure
4. Observation to study attitudes should be used when
 - A) the subject may not provide accurate information
 - B) the focus is directly on observable behavior
 - C) the evidence suggests that observation is more objective
 - *D) all of these
5. In the peer rating method, the respondent
 - A) rates each of the children in a classroom
 - *B) indicates the name of the child that fits a description
 - C) indicates which child they prefer
 - D) rates friends on intelligence, friendliness, etc.
6. In the _____ the median values for each item are computed.
 - A) method of summated ratings
 - B) Bogardus method
 - C) Guttman scaling
 - *D) method of equal appearing intervals
7. The method of equal appearing intervals ordinarily asks judges to sort items into _____ categories.
 - A) 5
 - B) 7
 - *C) 11
 - D) 15

8. In the Thurstone method the judges are usually
 - A) legal experts
 - *B) experts on the target being assessed
 - C) any captive volunteers
 - D) psychologists knowledgeable about scaling
9. In the Thurstone method, the _____ is used as a measure of variability:
 - A) standard deviation
 - *B) interquartile range
 - C) average deviation
 - D) pile number
10. A small Q value for an item in a Thurstone scale indicates that the item
 - A) is highly valid
 - B) is very unreliable
 - *C) shows substantial agreement as to placement
 - D) shows substantial agreement among subjects
11. The attitudes of the judges who do the initial sorting in the Thurstone procedure:
 - *A) do not affect the sorting, if careful instructions are given
 - B) affect the sorting significantly, but can be taken into account
 - C) can be altered prior to their task
 - D) can be counterbalanced statistically
12. A criticism of Thurstone and other scales:
 - A) the test-retest reliability cannot be determined
 - B) variations from the original procedure result in major differences
 - C) the interquartile range only looks at the middle of the distribution
 - *D) the same total score can be obtained by endorsing different items
13. The Liberalism-Conservatism scale discussed as an example of the Thurstone method, used _____ as judges.
 - *A) college students
 - B) Republican and Democratic registered voters
 - C) professional politicians
 - D) political science professors
14. Both the Thurstone and the Likert methods share these steps:
 - A) using expert judges and calculating medians
 - *B) choosing a target, and generating a pool of items
 - C) computing interquartile range, and total scores
 - D) item versus total correlations, and Q
15. The Bogardus scale attempts to measure:
 - *A) social distance
 - B) racial attitudes
 - C) attitudes towards marriage
 - D) political liberalism
16. The coefficient of reproducibility applies to:
 - A) all attitude scales
 - B) the method of summated ratings
 - C) Thurstone scales
 - *D) Guttman scales

17. In the index of reproducibility we need to compute the number of errors. Errors are here defined as:
 - A) endorsing extreme responses like “highly disagree”
 - *B) response patterns that show reversals
 - C) response patterns that do not follow majority consensus
 - D) inappropriate responses given by subjects
18. The Guttman methodology:
 - A) results in ratio scales
 - *B) has had an impact on the theory of scale construction
 - C) has resulted in many useful scales
 - D) yields a racial distance quotient
19. The major factors of the Semantic Differential are:
 - A) denotation, connotation, evaluation
 - B) liberalism, conservatism, optimism
 - C) agreement, disagreement, don’t know
 - *D) evaluation, potency, activity
20. A series of bipolar adjectives:
 - A) scalogram analysis
 - B) Survey of Interpersonal Values
 - *C) semantic differential
 - D) sociometric technique
21. The items in the Semantic Differential are selected on the basis of:
 - A) interquartile range and median
 - B) high coefficient of reproducibility
 - *C) factor representativeness and relevance
 - D) distance-cluster analysis
22. Each semantic differential item is usually scored
 - A) on a 5 point scale
 - B) zero or one
 - C) using median values
 - *D) on a 7 point continuum
23. The D statistic is a measure of
 - *A) distance between two concepts
 - B) meaning assigned to concepts
 - C) variability among judges
 - D) average endorsement of item
24. On the Semantic Differential the evidence suggests that
 - A) all bipolar adjectives are indeed bipolar
 - B) it doesn’t make any difference whether adjectives are bipolar
 - C) none of the standard adjectives are bipolar
 - *D) for some adjectives the assumption of bipolarity is not met
25. The Conservatism or C scale discussed in the text is a good example of:
 - A) a Semantic Differential Scale
 - *B) a check list
 - C) scalogram analysis
 - D) the method of summated ratings
26. The major scale to measure rigidity or dogmatism:
 - A) the C scale
 - B) the Bogardus scale
 - *C) the F scale
 - D) the rigid-dogmatic scale

27. Religious fundamentalism, resistance to change, and insistence on strict rules all refer to:
- A) items on the Rokeach
 - *B) conservatism
 - C) values
 - D) general occupational themes
28. Validity evidence for the Wilson-Patterson C scale includes:
- A) substantial correlations with Occupational scales of the Strong
 - B) item discrimination indices between retarded and normal children
 - C) very large significant cross-cultural differences
 - *D) gender differences, with females scoring higher
29. A criticism of the C scale:
- A) the evaluative factor is not very strong
 - *B) not unidimensional
 - C) the items are difficult to understand
 - D) poor cross-cultural stability
30. Numerical rating scales:
- *A) may have the numbers omitted from the actual form
 - B) always has the numbers printed that are assigned to each choice
 - C) ask the subject to select from a range of numbers
 - D) require the respondent to indicate 1 or 0 for each item
31. A(n) _____ scale uses a line or similar device where the respondent can place their answer.
- A) numerical
 - B) self-anchoring
 - *C) graphic
 - D) attitude scale
32. The self-anchoring scale provides a self-defined continuum of:
- A) attitudes towards religion
 - *B) personal perception
 - C) social distance
 - D) sociometric preference
33. In writing attitude scale items one should use
- A) "apple pie and motherhood" types of statements
 - B) emotionally laden words
 - C) factual statements
 - *D) positive rather than negative wording
34. The Study of Values was based on:
- *A) Spranger's theory of six value types
 - B) Guttman's scalogram analysis
 - C) Rokeach's distinction between two types of values
 - D) Gordon's theory of basic motivational patterns
35. A unique aspect of the Study of Values:
- A) the scales were correlated positively
 - *B) it could be hand scored by the subject
 - C) there was little evidence of its validity
 - D) scores were not related to different occupations
36. The text indicates that "as expected" the scales of the Study of Values correlate negatively with each other; this is because:
- A) Spranger did not consider the possibility of a "valueless" person
 - B) the reliability of the scales is rather low
 - *C) in an ipsative instrument high scores on one scale mean low scores on the other

37. In the study by Huntley (1965) of what happens to the values of male college students between entrance and graduation, it was found that:
- *A) aesthetic values increase regardless of major
 - B) all values increased substantially
 - C) only science majors changed their values substantially
 - D) both social and political values increased dramatically
38. This instrument looks at instrumental and terminal values:
- A) The Study of Values
 - *B) The Rokeach Value Survey
 - C) The Survey of Interpersonal Values
 - D) The IT study of values
39. According to Rokeach values, as opposed to attitudes
- A) are much smaller in number
 - B) occupy a more central position
 - C) refer to a single belief
 - *D) all of these
40. The items that make up the Rokeach Value Survey were arrived at basically by
- A) criterion-keying
 - *B) the FIAT method
 - C) factor analysis
 - D) the judgment of experts
41. A basic question regarding the Rokeach is that of _____ validity.
- *A) content
 - B) construct
 - C) concurrent
 - D) tertiary
42. Once the subject gives the required ranks on the Rokeach, they can then be
- A) changed to z scores
 - *B) compared to someone else's ranks
 - C) summed for a total score
 - D) correlated with each other
43. If we wanted to compare men versus women on the Rokeach, we could
- *A) summate subsets of values together
 - B) add the ranks to get a total
 - C) compute medians and compare these
 - D) assess the temporal stability of the items
44. In regards to the reliability of the Rokeach:
- A) excellent alternate form reliability
 - B) adequate test-retest reliability
 - C) excellent only for terminal values
 - *D) it is marginal at best
45. The construct validity of the Rokeach:
- *A) seems well supported by the literature
 - B) is probably its weakest link
 - C) has not been investigated
 - D) based only on U.S. studies
46. In the study by Domino and Acosta (1987) of first generation Mexican Americans, it was found that:
- A) Mexican Americans had distinctly different values from control subjects
 - B) there was no difference in values between groups
 - *C) subjects who were more American in language were more American in values
 - D) the less accultured subjects endorsed more terminal values

47. From a factor analytic point of view, on the Rokeach:
- A) there seems to be one general factor similar to “g”
 - *B) the factors on the Rokeach are not very powerful
 - C) the Rokeach seems like a very strong instrument
 - D) most of the items are included in one of the factors
48. In the normative value rankings presented by Rokeach:
- A) there were no significant gender differences
 - B) the ranks given by children are substantially different from adults
 - C) scores on the leadership scale correlate significantly with the total
 - *D) the gender differences support the construct validity
49. On the Rokeach, gender differences:
- *A) males endorse more achievement, materialism, and pleasure seeking
 - B) females endorse more intellectually oriented values
 - C) females rank personal happiness lower than males
 - D) do not match how men and women are socialized in Western cultures
50. In the formula for the rank order correlation coefficient, N stands for:
- A) the number of subjects in one’s experiment
 - B) how much longer one scale is than another
 - *C) the number of items being ranked
 - D) how many items are needed to achieve .70
51. The Rokeach is a(n) _____ scale.
- A) nominal
 - *B) ordinal
 - C) interval
 - D) ratio
52. A major criticism of the Rokeach is that:
- A) the norms are limited to college students
 - B) rating each value would be a better procedure
 - C) the literature does not support its construct validity
 - *D) the ranking does not allow the measurement of intensity
53. Both the Survey of Interpersonal Values and its companion use:
- A) lists of values
 - *B) forced choice triads
 - C) true-false items
 - D) multiple choice items
54. The six scales of the Survey of Interpersonal Values:
- *A) were based on factor analysis
 - B) reflect Gordon’s theory about values
 - C) were chosen to summarize the 18 Rokeach values
 - D) reflect major career interests
55. Items used in the Survey of Interpersonal Values were originally matched on social desirability. This procedure:
- A) worked quite well, so the SIV scales do not correlate with social desirability
 - B) did not need to be carried out since values are not influenced by social desirability
 - *C) apparently did not eliminate social desirability as a confound
 - D) doesn’t make sense given the type of items used in the SIV
56. The scales of the Survey of Interpersonal Values
- *A) correlate negatively with each other, since this is an ipsative scale
 - B) correlate positively with each other, since this is an ipsative scale
 - C) do not correlate with each other since factor analysis was used
 - D) intercorrelate positively because the items are triads

57. SIV scales do not correlate with measures of intelligence. This might be considered evidence for _____ validity.
- A) primary
 - *B) secondary
 - C) tertiary
 - D) factorial
58. Gordon, the author of the SIV, argues that the SIV scales
- *A) have factorial validity
 - B) show significant gender differences
 - C) are highly predictive of occupational status
 - D) are a good example of Guttman scales
59. In regards to the validity of the SIV:
- *A) the manual gives considerable data
 - B) the literature does not support its validity
 - C) it is marginal at best
 - D) needs to be investigated
60. A major issue with the Survey of Interpersonal Values is that:
- A) scores correlate significantly with measures of intelligence
 - B) the scaled values are not important in Western cultures
 - C) the true-false format yields little information
 - *D) the scaled dimensions are personality variables
61. A good source of information about the Survey of Interpersonal Values:
- A) the internet
 - *B) the Mental Measurements Yearbook
 - C) the book "Human Values"
 - D) the Annual Review of Psychology
62. A major concern with most instruments of values:
- *A) content validity
 - B) concurrent validity
 - C) face validity
 - D) primary validity
63. If we ask a person what they are interested in, their response is considered:
- A) conscious interests
 - B) invalid interests
 - *C) expressed interests
 - D) inventoried interests
64. A major difference between the original Strong and Kuder inventories:
- A) the Strong used content scales, the Kuder basic interest scales
 - B) the Strong used general occupational themes, the Kuder specific themes
 - C) the Strong used Holland's theory, the Kuder did not
 - *D) the Strong used occupational scales, the Kuder content scales
65. The link between career interests and personality was primarily made by:
- A) Strong
 - B) Kuder
 - *C) Holland
 - D) Campbell

66. The text states that a “third key event in the history of career interest assessment” concerned a personality theory. What were the first two events?
- *A) the development of the Strong and of the Kuder
 - B) Wundt’s laboratory and Binet’s test
 - C) the Binet-Simon and Wechsler-Bellevue scales
 - D) none of these
67. There is little support for the idea that:
- A) people with similar interests tend to enter the same occupation
 - *B) job satisfaction is a function of similarity of interests
 - C) people with different interests tend to enter similar occupations
 - D) on the Strong, item content is more important than item format
68. Basically, the Strong compares a person’s career interests with those of people
- *A) satisfactorily employed in a variety of occupations
 - B) with the same personality type
 - C) who are considering entering a particular occupation
 - D) who are outstanding examples of various occupations
69. The Strong uses different types of items because
- A) research has shown that item format is more important than item content
 - *B) it would be too boring to respond to one kind of item only
 - C) different weights are given to different types of items
 - D) different types of items yield different variables
70. Which of these scales would not likely be found on the Strong?
- A) biologist
 - B) engineer
 - *C) plumber
 - D) artist
71. The Strong is very useful in:
- A) identifying potential careers that a person should enter
 - B) identifying the talent a person has for a specific career
 - C) helping high school dropouts find a job
 - *D) none of these
72. The item “college professor” is a good item for the Strong because
- *A) it has a large response range
 - B) it has a narrow response range
 - C) most people like that occupation
 - D) most people dislike that occupation
73. Items on the Strong were retained on the basis of:
- *A) multiple criteria
 - B) their “public relations” aspects
 - C) their predictive validity
 - D) their gender appropriateness
74. Both the Strong and the CPI are alike in that
- A) they use true-false items
 - B) they share the same theoretical framework
 - C) they used the fiat method of scale development
 - *D) they are open systems

75. In selecting an occupational group for a potential Strong scale, which criteria does NOT apply:
- A) between the ages of 25 and 55 or 60
 - B) perform work that is typical of that occupation
 - C) satisfied in their occupation
 - *D) be a member of the pertinent union or organization
76. In regards to the various criterion groups used for the Strong, they are typically:
- *A) large (i.e., about 200 to 400)
 - B) small (usually less than 100)
 - C) composed of both men and women
 - D) composed of retired individuals
77. The comparison group for the Strong:
- A) consisted of 200 men that Strong had originally tested
 - *B) has taken different forms over the years
 - C) depends upon the specific occupation
 - D) varies from scale to scale
78. The Strong
- *A) is returned to the publisher or other company for scoring
 - B) can be scored by hand
 - C) does not have a formal scoring system
 - D) can be hand scored only by a trained technician
79. The Administrative Indices on the Strong:
- A) indicate whether the subject would be a good administrator
 - B) compares the person's interests with various administrative jobs
 - *C) are routine clerical checks performed by the computer
 - D) are ways of assessing the reliability of the test
80. A set of six scales designed to portray a "general" type (like realistic or artistic) on the Strong:
- A) Basic Interest scales
 - B) Occupational scales
 - C) Administrative scales
 - *D) General Occupational themes
81. On the SVIB, occupational areas like agriculture or sales are considered:
- A) Administrative indices
 - *B) Basic interest scales
 - C) General occupational themes
 - D) Special scales
82. On the SVIB, the Occupational scales are typically made up of:
- *A) heterogeneous items
 - B) homogeneous items
 - C) items that match Holland's theory
 - D) items that have substantial face validity
83. The Academic Comfort Scale on the Strong attempts to differentiate between
- A) those who do well academically and those who don't
 - B) college professors and research scientists
 - *C) those who enjoy being in an academic setting and those who do not
 - D) those who should continue to graduate school and those who should not

84. The criterion used in developing the Introversion-Extroversion scale of the Strong:
- A) self ratings of introversion or extroversion
 - *B) scores on the MMPI Introversion-Extroversion scale
 - C) whether the person prefers working with people or things
 - D) psychologists' judgment of how introverted a person is
85. The scores on the Strong are primarily presented as:
- A) z scores
 - *B) T scores
 - C) lambda scores
 - D) percentiles
86. Interpreting and communicating the results of the Strong to a client
- A) is a relatively routine procedure
 - B) does not require an examiner
 - C) should be done only by a Ph. D. level psychologist
 - *D) requires a sophisticated and well trained examiner
87. Individuals who are aggressive and prefer concrete to abstract activities, should score high on the _____ occupational theme.
- A) artistic
 - B) aggressive
 - *C) realistic
 - D) enterprising
88. Occupations like credit manager, accountant, and secretary:
- A) investigative
 - *B) conventional
 - C) enterprising
 - D) social
89. In regards to the reliability of the Strong:
- A) good test-retest but only with short time periods
 - *B) substantial test-retest reliability even with long time periods
 - C) excellent alternate form reliability, especially in the latest revision
 - D) somewhat poor concurrent reliability
90. If one studies the predictive validity of the Strong, the criterion is typically:
- A) the major a person has in college
 - B) the total score on a specific occupational scale
 - *C) the occupation the person eventually enters
 - D) one's type as defined by the General Occupational Themes
91. In regards to faking on the Strong, we can probably conclude that:
- *A) it's possible but does not usually occur
 - B) it is not possible because of the Administrative Indices
 - C) it occurs to a substantial degree and results in invalid profiles
 - D) it has not been of concern to investigators
92. The statistic of percent overlap:
- *A) percentage of scores in one sample matched by scores on the second
 - B) difference between 25th and 75th percentiles
 - C) a way of assessing faking on the Strong
 - D) percent of females who score higher than males
93. Percent overlap is a way of expressing _____ validity.
- A) factorial
 - B) discriminant
 - C) tertiary
 - *D) concurrent

94. Racial differences on the Strong:
- A) are pronounced especially with Hispanics
 - *B) have not been studied extensively
 - C) occur primarily on the Basic Interest scales
 - D) are present on the General Occupational Themes only
95. The item response distribution on the Strong
- *A) is simply the endorsement of an item given by different groups
 - B) assesses the reliability of individual items
 - C) can be computed only for the Basic Interest Scales
 - D) tells how a particular subject responded
96. In regards to the Strong, from a longitudinal aspect:
- A) scales developed in the 1930's are no longer valid
 - *B) revised Strong scales are usually fairly close to the original
 - C) an occupational scale will only be valid for about 5 years
 - D) all of these
97. The Kuder was originally developed on the basis of:
- A) criterion group scaling
 - *B) homogeneous scaling
 - C) criterion-keying
 - D) longitudinal studies
98. Unlike the Strong, the Kuder Occupational Interest Survey also has:
- A) administrative indices
 - B) occupational scales
 - *C) college major scales
 - D) a social desirability scale
99. The scoring procedure on the Kuder:
- *A) reflects most and least preferred choices
 - B) is basically identical to that of the Strong
 - C) gives equal weights to true and to false
 - D) is rather simple and can be done by hand
100. Although the Strong and the Kuder use different scoring procedures, the end result is that both:
- A) use Holland's theory to generate Basic Interest Scales
 - B) are excellent predictors of success in an occupation
 - C) show relatively poor predictive validity
 - *D) focus on the similarity of interest between a person and a criterion group
101. Interest inventories for disadvantaged clients:
- A) use identical items as on the Strong
 - B) have excellent reliability and validity
 - *C) often use drawings of people or activities
 - D) cover only occupations that require a college degree
102. Most career interest inventories are designed for occupations
- A) that would be labelled as "blue collar"
 - *B) entered by middle-class individuals
 - C) that are of interest primarily to females
 - D) that involve the military and government service

103. If you were counseling normal individuals most likely to enter nonprofessional occupations, you might want to use the:
- A) Strong Vocational Interest Blank for Men
 - B) Strong Interest Inventory
 - C) Geist Picture Interest Inventory
 - *D) Career Assessment Inventory
104. One criticism of the field of career interest measurement:
- A) most instruments do not work well
 - *B) dominated by an empirical approach
 - C) too biologically and genetically oriented
 - D) strongly focused on theory and construct validity
105. Should we continue to develop new occupational scales?
- A) Yes; new occupations are developing all the time
 - B) Yes; the old scales are becoming rapidly outdated
 - *C) No; any new occupation can be located in current theoretical frameworks
 - D) No; the whole approach does not work very well

7 Psychopathology

1. Physical illness is to mental illness as:
 - A) the MMPI is to the Mental Status Exam
 - *B) International Classification of Diseases is to the DSM
 - C) MMY is to the Diagnostic and Statistical Manual
 - D) MMPI-2 is to the STAI
2. The DSM is basically
 - *A) a guide intended to assist in making diagnoses
 - B) like a cookbook where a collection of symptoms equals a disease
 - C) a dictionary of mental illness terms and types
 - D) a listing of all known physical and mental illnesses
3. A mental status exam would typically cover all of these except:
 - A) intellectual functioning
 - B) thought processes
 - *C) family background
 - D) appearance
4. The mental status exam:
 - A) has very poor reliability
 - B) consists of multiple choice items
 - C) used primarily for differential diagnosis
 - *D) is not well standardized
5. The SCID can be described as a
 - *A) semistructured interview to assess psychopathology
 - B) newer version of the Diagnostic and Statistical Manual
 - C) revision of the MMPI
 - D) a scoring procedure for the MCMI
6. A self-report screening inventory for psychopathology:
 - A) SCID
 - B) STAI
 - C) MCMI
 - *D) SCL-90R
7. The nine basic dimensions of the SCL-90R were originally developed using:
 - A) factor analysis exclusively
 - B) criterion-keying
 - *C) both factor analysis and rational judgement
 - D) contrasted groups

8. On the SCL-90R the Global Severity Index:
 - A) is a measure of intensity of checked symptoms
 - *B) reflects both the number of symptoms and their intensity
 - C) is simply the number of items checked
 - D) is equivalent to the K scale of the MMPI
9. In general, the reliability of the SCL-90R:
 - *A) is quite adequate
 - B) is only of the internal consistency type
 - C) does not apply to checklists
 - D) has not been investigated
10. We would expect concurrent validity of the SCL-90R to include correlations with:
 - A) the Beck Depression Inventory
 - *B) scales on the MMPI
 - C) the State-Trait Anxiety Inventory
 - D) future suicide attempts
11. Factorial invariance means:
 - A) do the factors show a narrow range of scores
 - B) is there substantial variation in scores from group to group
 - *C) is the factor structure the same in different groups
 - D) do such tests as the MMPI and the SCL-90R have the same factors
12. The Alienation scale of the PSI attempts to identify:
 - *A) those who respond like hospitalized psychiatric patients
 - B) individuals who feel alienated from society
 - C) people who are potentially at risk for suicide
 - D) those who respond to the PSI in a socially undesirable manner
13. The social non-conformity scale of the PSI was developed by comparing the responses of
 - A) normals versus schizophrenics
 - B) conforming and nonconforming college students
 - *C) normals versus reformatory inmates
 - D) the same individuals tested twice
14. The Expression Scale of the PSI is essentially a measure of:
 - *A) extroversion and control
 - B) psychotic thinking
 - C) defensiveness
 - D) sociopathy
15. On most instruments like the MMPI, scores that are within _____ are considered normal.
 - A) twenty points of the mean
 - *B) two SDs from the mean
 - C) T scores of 50 and 80
 - D) one SD above or below the mean
16. The reliability of the PSI suggests that:
 - A) the test is stable over time but has little internal consistency
 - B) the test has excellent internal consistency
 - *C) great care should be used in making decisions about individuals
 - D) alternate forms reliability is marginal
17. In the validation study of the PSI where suicidal versus non-suicidal students were assessed:
 - A) suicidal students scored lower on most PSI scales
 - *B) suicidal students scored higher on most PSI scales
 - C) there were no significant differences between the two groups
 - D) the results supported the discriminant validity of the PSI

18. In the Overall (1974) study of the PSI comparing the original scales versus those obtained from factor analysis:
 - *A) the original scales worked somewhat better
 - B) the factor scales were significantly better
 - C) both sets of scales worked equally well
 - D) both sets of scales were criticized as invalid
19. Among the criticisms of the PSI:
 - A) scales intercorrelate
 - B) may be affected by social desirability
 - C) the names of the scales are not the best
 - *D) all of these
20. Prior to the MMPI most personality inventories were developed using:
 - *A) the logical keying approach
 - B) the empirical keying approach
 - C) contrasted groups
 - D) factor analysis
21. MMPI scales were formed by identifying subgroups of items that differentiated between:
 - A) mentally ill and normals
 - *B) a specific diagnostic group, normal subjects, and other diagnostic groups
 - C) psychiatric patients with and without a diagnosis
 - D) homogeneous samples of hospital visitors
22. The standard MMPI profile includes eight psychiatric scales and a(n)
 - *A) MF scale and a Social Introversion scale
 - B) Masculinity scale and a Femininity scale
 - C) Hypomania scale and a Depression scale
 - D) Depression scale and an Anxiety scale
23. The Lie scale on the MMPI is composed of items that:
 - A) have high correlations with social desirability
 - B) are quite heterogeneous but cover poor physical health
 - C) showed a significant response shift between lie and honest instructions
 - *D) most people, if answering honestly, would not endorse
24. The K scale of the MMPI was developed by comparing psychiatric patients and normals:
 - A) both with abnormal profiles
 - B) both told to lie
 - *C) both with normal profiles
 - D) who were related by marriage
25. The F scale on the MMPI is composed of items that:
 - A) most honest people endorse as true
 - *B) fewer than 10% of the normative samples endorsed in a particular way
 - C) showed a significant response shift under instructions to lie
 - D) differentiated normal from psychiatric samples
26. Originally, the MMPI was validated against the criterion of:
 - *A) psychiatric nosology
 - B) DSM-III
 - C) normality
 - D) medical illness

27. One of the major concerns of the original MMPI:
 - A) the psychiatric diagnoses were not valid
 - B) the inventory was too long
 - *C) the standardization sample was not representative
 - D) the original sample did not include children
28. The MMPI 2:
 - *A) is highly similar to the original MMPI
 - B) can for all practical purposes be considered a new test
 - C) contains an entirely new set of scales
 - D) is aimed at adolescents rather than adults
29. The MMPI 2 contains three new scales:
 - A) they are clinical scales of anxiety and depression
 - B) to cover new psychiatric conditions like post-traumatic stress
 - *C) they are validity rather than clinical scales
 - D) that measure drug abuse and suicide potential
30. The T scores on the MMPI 2:
 - A) are standard T scores with mean of 50
 - *B) take into account the nonnormality of the raw scores
 - C) have to be computed by the examiner
 - D) are identical to percentiles
31. In the development of the clinical MMPI scales, the primary focus was on _____ validity.
 - A) primary
 - B) construct
 - C) concurrent
 - *D) empirical
32. Harris and Lingoes developed subscales on the MMPI on the basis of:
 - *A) clinical judgment
 - B) factor analysis
 - C) psychiatric diagnosis
 - D) criterion-keying
33. Content scales for the MMPI 2:
 - A) are identical to the original clinical scales
 - B) were developed on the basis of factor analysis
 - *C) used both expert judgment and internal consistency
 - D) generally yield two basic clusters
34. On the MMPI, there are at least two ways of conducting a factor analysis. We can factor analyze the _____, or we can factor analyze the _____.
 - A) subjects; clinical scales
 - B) clinical scales; content scales
 - *C) scale scores; responses to the items
 - D) total score; scores on the clinical scales
35. On the MMPI, homogeneous scales developed by factor analysis:
 - A) cannot be developed because MMPI items are heterogeneous
 - *B) have been developed but are not widely used
 - C) are not useful since they don't follow the DSM
 - D) have been developed and are regularly used
36. The Ego strength scale was developed to predict:
 - A) alcoholism
 - B) five aspects of psychopathology
 - C) inner religious belief
 - *D) success in psychotherapy

37. The MacAndrew Alcoholism scale was developed by comparing the responses of:
- *A) alcoholics versus psychiatric patients
 - B) alcoholics versus normals
 - C) psychiatric patients with and without depression
 - D) alcoholic patients before and after treatment
38. In general, on the MMPI we would expect _____ to be higher than _____.
- A) internal consistency; alpha coefficients
 - B) the reliability of configural patterns; the reliability of individual scales
 - C) internal consistency; test-retest reliability for a brief period
 - *D) test-retest reliability for a brief period; alpha coefficients
39. In regards to the validity of the MMPI we can conclude that:
- A) the validity of the MMPI is quite solid and basically a closed issue
 - *B) the issue is a complex one, but there is some support for validity
 - C) the validity of the MMPI is rather poor from any point of view
 - D) from an empirical point of view, the validity is excellent
40. In regards to racial differences on the MMPI, the differences
- A) are major, especially with black subjects
 - *B) appear to be minor, but caution needs to be exercised
 - C) have not been investigated to any degree
 - D) can be simply ignored as error variance
41. Among the criticisms aimed at both the MMPI and the MMPI 2, which is NOT correct:
- A) lacks a theoretical model
 - B) the scales are heterogeneous
 - *C) lacks content validity
 - D) response style may play a role
42. This test was developed as a better and more modern version of the MMPI:
- *A) Millon Clinical Multiaxial Inventory
 - B) State-Trait Anxiety Inventory
 - C) Center for Epidemiologic Studies Scale
 - D) Wisconsin Personality Disorders Inventory
43. As contrasted to the MMPI, the MCMI:
- *A) has a well defined theoretical rationale
 - B) is substantially longer than the MMPI
 - C) does not use the DSM for diagnostic criteria
 - D) can be administered to a group
44. The computer generated narrative report on the MCMI:
- A) provides a lay person's summary to be shared with the patient
 - B) is designed to be read by the patient and his/her family
 - C) provides a summary of the client's personality
 - *D) is considered for professional use only
45. The first step in the development of the MCMI was to:
- A) identify patients that represented pure clinical syndromes
 - B) develop two forms of the inventory
 - *C) create a pool of self-descriptive items
 - D) locate items that correlated with MMPI scales
46. We can criticize Millon's procedures in developing the MCMI because:
- A) his item pool was too small
 - *B) the items were keyed true
 - C) he did not address content validity
 - D) test-retest reliability was not considered

47. In the second phase of the development of the MCMI, items were retained:
- *A) because they correlated with the total score on the designated scale
 - B) if they showed extreme endorsement frequencies
 - C) that significantly correlated with MMPI scale scores
 - D) which agreed with Millon's theory about psychopathology
48. On the MCMI, each item
- A) appears on only one scale
 - B) is scored on every scale, though the keyed response may differ from scale to scale
 - C) may be scored on more than one scale, though the keyed response is always the same
 - *D) typically is scored on about 4 scales, though the keyed response may differ from scale to scale
49. In Millon's theory, patients who evaluate satisfaction in terms of the reaction of others are said to be _____ types.
- A) ambivalent
 - B) detached
 - *C) dependent
 - D) passive
50. Millon's theory about personality disorders is based on two dimensions:
- A) coping and independence
 - *B) reinforcement and coping
 - C) activity and passivity
 - D) detachment and attachment
51. One advantage of the MCMI over the MMPI is that:
- A) there are parallel forms of the MCMI
 - B) the MCMI has been around much longer
 - C) the MCMI can be easily hand scored
 - *D) its scales more closely parallel the DSM
52. The validity index on the MCMI
- *A) is composed of 4 items endorsed by less than 1% of patients
 - B) parallels the K scale of the MMPI
 - C) is composed of items that discriminate normals from psychiatric patients
 - D) assesses the degree to which the person fakes bad
53. One advantage of computer scoring over hand scoring is that:
- *A) refinements and/or changes in scoring can be easily made on the computer
 - B) each examiner can score a particular test protocol
 - C) scorer reliability can be more easily assessed
 - D) computer scoring is more widely available
54. The MCMI uses
- A) regular T scores
 - B) uniform T scores
 - *C) base rate scores
 - D) integrated z scores
55. The problem with the scoring procedure on the MCMI is that:
- A) the T scores used do not take account of nonnormality of scores
 - *B) scores reflect the base rate of the standardization sample
 - C) different patients receive different scores
 - D) it is impressionistic rather than statistical

56. Which type of scale on the MCMI would have the highest reliability?
- *A) personality pattern
 - B) pathological personality
 - C) clinical syndromes
 - D) validity scales
57. The MCMI illustrates well the argument that
- A) validity should be the last step to be checked
 - B) primary validity is needed before anything else
 - *C) validity should be incorporated in all phases of test construction
 - D) reliability and validity cannot be easily separated
58. In general, the validity of the MCMI:
- A) has not been adequately explored
 - *B) seems supportive of the instrument's usefulness
 - C) is adequate but only from a content perspective
 - D) has not at all been criticized
59. The results of factor analyses of the MCMI
- *A) are in agreement with clinical theory and experience
 - B) clearly support the independence of the MCMI scales
 - C) suggest a factorial invariance across different groups
 - D) support the idea of one general factor of maladjustment
60. Both the MCMI and the Wisconsin Personality Disorders Inventory:
- A) evolved from the MMPI
 - B) reflect Millon's theory
 - C) have scales that are statistically independent
 - *D) are based on the DSM
61. A narcissistic individual can be characterized as:
- A) having emotional blandness
 - B) submissive
 - *C) having an inflated self-image
 - D) aggressive
62. High scorers on the MCMI Histrionic scale can be characterized as:
- *A) immature and childish
 - B) self sacrificing and masochistic
 - C) isolated and empty
 - D) verbally and physically hostile
63. The schizoid scale and the schizotypal scale on the MCMI differ in that the schizoid scale reflects _____, and the schizotypal scale reflects _____.
- A) brief lasting symptoms; everyday functioning
 - *B) everyday functioning; chronic pathology
 - C) normal behavior; abnormal behavior
 - D) emotional blandness; anxiety
64. On the Wisconsin Personality Disorders Inventory, test items are answered:
- *A) on a 10 point rating scale
 - B) true or false
 - C) using a Likert response scale
 - D) by selecting "most" and "least" responses
65. The norms on the Wisconsin Personality Disorders Inventory:
- A) represent the major personality syndromes
 - B) were carefully selected through use of census data
 - C) are based on hospital samples in the U.S. and England
 - *D) are clearly samples of convenience

66. On the Wisconsin Personality Disorders Inventory, reliability was computed using alpha coefficients. This would alert us to the fact that:
- A) the items are true-false
 - *B) different response options get different weights
 - C) there are no alternate forms
 - D) the scales are made up of heterogeneous items
67. One major problem with the Wisconsin Personality Disorders Inventory:
- *A) the scales correlate substantially with each other
 - B) the scales do not correlate with parallel scales on the MCMI
 - C) the scaled variables are of limited usefulness
 - D) the normative samples were too small
68. The Schizotypal Personality Questionnaire was developed using the method of:
- A) fiat
 - *B) internal consistency
 - C) content analysis
 - D) factor analysis
69. The original sample on which the Schizotypal Personality Questionnaire was developed consisted of:
- *A) volunteer college students
 - B) Counseling Center clients
 - C) psychiatric patients
 - D) normal but eccentric individuals
70. In the development of the Schizotypal Personality Questionnaire "items were deleted if the item-total correlation was less than .15". This was done to:
- A) shorten the test to a reasonable length
 - B) select only items with high test-retest reliability
 - C) assure the construct validity of the test
 - *D) develop more homogeneous subscales
71. In assessing the convergent and discriminant validity of the Schizotypal Personality Questionnaire, "discriminant validity coefficients were .19 and .37"; this shows that:
- A) the test is not valid since it does not discriminate well
 - *B) the discriminant validity is low as it should be
 - C) the convergent validity must be high
 - D) scores on this test correlate significantly with parallel scores on other tests
72. A transitory emotional state characterized by subjective feelings of tension.
- *A) state anxiety
 - B) schizotypal personality
 - C) trait anxiety
 - D) narcissism
73. Scoring of the STAI:
- A) requires use of a computer
 - *B) is typically done by hand
 - C) is a complicated matter
 - D) uses base rate scores
74. Test-retest reliability coefficients for the state portion of the STAI are low because:
- A) trait anxiety is seen as transitory
 - *B) the instructions and treatments were designed to alter scores
 - C) the test has in fact low reliability
 - D) the state scale is not homogeneous

75. The concurrent validity of the STAI was shown with:
- *A) other measures of anxiety
 - B) psychiatric patients
 - C) factor analysis of the state portion
 - D) comparisons with the MMPI
76. STAI scores and academic achievement:
- A) are correlated substantially
 - B) show a significant correlation but only for state scores
 - C) are significantly related in males but not in females
 - *D) seem to be unrelated in college students
77. The factor analytic results on the Beck Anxiety Inventory indicate that:
- A) anxiety is made up of two primary factors
 - *B) the structure of anxiety differs in psychiatric patients and normals
 - C) the homogeneity of the factors is poor
 - D) the factor structure differs for males and females
78. To assess the discriminant validity of the Beck Anxiety Inventory, it would be “natural” to use:
- A) the MMPI scale of anxiety
 - B) several of the scales of the MCMI
 - *C) measures of depression
 - D) measures of intelligence
79. One of the unusual aspects of the Beck Depression Inventory is that it was developed:
- *A) by fiat
 - B) on college students
 - C) as part of the MMPI
 - D) originally as an anxiety scale
80. The items on the Beck Depression Inventory represent _____ symptoms of depression.
- A) somatic
 - B) affective
 - C) subtle
 - *D) cognitive
81. Initially the Beck Depression Inventory could be considered as a(n):
- *A) structured interview
 - B) mental status exam
 - C) projective technique
 - D) depression checklist
82. Scores on the BDI are:
- *A) raw scores
 - B) T scores
 - C) z scores
 - D) percentiles
83. The Beck Depression Inventory has a “floor effect.” This means that:
- A) high scores do not indicate happiness
 - B) the range of high scores is limited
 - *C) low scores do not necessarily indicate happiness
 - D) the scale measures both depression and happiness
84. This type of reliability is somewhat problematic for instruments like the Beck Depression Inventory.
- A) internal consistency
 - *B) test-retest
 - C) alternate forms
 - D) split-half

85. We would expect the test-retest reliability of the Beck Depression Inventory to be lower in _____ than in _____.
A) normals; psychiatric patients
B) males; females
C) adults; adolescents
*D) psychiatric samples; normals
86. The internal consistency of the BDI is typically in the:
A) .90s
*B) .80s
C) .70s
D) .60s
87. In general, the construct validity of the BDI
A) has not been investigated to any degree
B) seems rather marginal
*C) has been supported in a variety of studies
D) is definitely inadequate
88. The secondary validity of the BDI:
A) includes a significant negative correlation with age
*B) seems to be supported by a number of findings
C) suggests that there are no gender differences
D) has not been investigated to any extent
89. One criticism of the Beck Depression Inventory is that the response format
A) requires the subject to carefully consider all alternatives
B) is too time consuming for most clients
*C) is highly susceptible to a position response
D) uses forced choice and is therefore not “user friendly”
90. The Center for Epidemiologic Studies – Depression scale was included in the text to illustrate how:
*A) a scale can be constructed from the best items of other scales
B) a screening inventory can be most useful
C) a scale can be developed to assess DSM criteria
D) the reliability of a test can be increased
91. The test-retest reliability of the Center for Epidemiologic Studies – Depression scale is:
A) unexpectedly high for a brief instrument
*B) on the low side
C) substantially higher than that of the Beck
D) higher than its internal consistency reliability
92. In regards to a core battery of tests that could be used by practitioners to assess psychopathology, there is:
*A) disagreement as to which tests might be included
B) agreement that the Zung belongs in such a battery
C) empirical evidence to suggest that the BDI should not be included
D) evidence that such uniformity would be highly desirable
93. Focal assessment involves the use of:
A) broad based instruments like the MMPI
B) a broad based battery to be given to all patients
C) focus groups to discuss specific issues
*D) specialized instruments, such as a measure of depression

94. The base rate affects:
- A) sensitivity
 - B) specificity
 - *C) predictive power
 - D) reliability
95. In order to deal with base rates and how they affect the error rate, we need to:
- *A) calculate local base rates and pay attention to them
 - B) calculate the specificity and sensitivity rates
 - C) select a sample with equal numbers of males and females
 - D) use instruments like the CES-Depression scale
96. Loss of appetite, sleep disturbance, and pessimism would likely be categories or items of the:
- A) Schizotypal Personality Questionnaire
 - *B) Beck Depression Inventory
 - C) Ego strength scale of the MMPI
 - D) Hostility scale of the SCL-90 R

8 Normal Positive Functioning

1. Items on the Tennessee Self Concept Scale:
 - A) true-false
 - *B) 5 response options from completely false to completely true
 - C) multiple choice with one correct answer
 - D) forced choice triads
2. The two basic dimensions by which items on the Tennessee Self Concept Scale were initially selected and classified:
 - *A) internal versus external aspects
 - B) identity versus self-satisfaction
 - C) private self versus public self
 - D) self esteem versus self efficacy
3. The final items retained on the Tennessee Self Concept Scale were retained based on:
 - A) factor analysis
 - B) high test-retest reliability
 - *C) psychologists' judgment
 - D) substantial item-total correlations
4. The Tennessee Self Concept Scale also contains items from the:
 - A) SCL-90
 - B) Kaufman's Self Concept Scale
 - C) Mental status exam
 - *D) MMPI Lie Scale
5. The Tennessee Self Concept Scale might be described as:
 - *A) self-administered, group test, with no time limit
 - B) requires an experienced examiner, has a time limit
 - C) extremely long to administer, individual test
 - D) individual timed test requiring college level reading skills
6. What is unusual about the two forms of the Tennessee Self Concept Scale?
 - A) one form is for the real self, the other for the ideal self
 - *B) the two forms are identical
 - C) the instructions on the two forms differ
 - D) one form is for males, the other for females
7. Hand scoring of the Tennessee Self Concept Scale
 - A) is not possible and must be machine scored
 - B) can be easily carried out by the client
 - *C) can be quite time consuming
 - D) is easy since only a total score is involved

Normal Positive Functioning

8. The original items on the Tennessee Self Concept Scale showed interjudge agreement. This is incorrectly interpreted by the test author as:
 - A) construct validity
 - B) test-retest reliability
 - C) inter-rater reliability
 - *D) content validity
9. The criterion validity data presented for the Tennessee Self Concept Scale:
 - A) supports its validity as a measure of self-concept
 - B) does not support its validity
 - C) is limited to studies of captive college students
 - *D) supports its validity as a measure of a global variable like adjustment
10. In regards to norms, the Tennessee Self Concept Scale:
 - *A) appears to be based on a sample of convenience
 - B) has excellent norms listed by age, gender, and other aspects
 - C) no explicit norms are given other than results of studies
 - D) because it is an ipsative measure, norms are not needed
11. In general, the Tennessee Self Concept Scale can be characterized as:
 - A) the best available measure of self-concept
 - *B) a weak instrument
 - C) having excellent reliability and validity
 - D) useful for both research and counseling purposes
12. The Self-Esteem Questionnaire:
 - A) shows excellent reliability
 - *B) has somewhat low internal consistency
 - C) shows poor test-retest reliability
 - D) is particularly useful because of its alternate forms
13. Primary validity on the Self-Esteem Questionnaire was shown by:
 - A) the items coming from a humanistic framework
 - B) no significant gender differences
 - *C) significant mean differences on six student groups
 - D) Q sort descriptions that were internally consistent
14. An example of tertiary validity for the Self-Esteem Questionnaire:
 - A) a factor analysis suggested one major factor
 - *B) students who produced creative achievements scored higher
 - C) scores on the test correlated significantly with other self-concept measures
 - D) ACL descriptions characterized high scorers as confident and productive
15. Internal control means:
 - A) the person is rigid and opinionated
 - B) the person has low self-esteem
 - C) the perception that one's behavior is controlled by powerful others
 - *D) the perception that rewards are contingent upon one's behavior
16. The items on the Internal-External Locus of Control scale are presented in forced choice pairs:
 - *A) one internal and one external
 - B) matched in social desirability
 - C) one masculine and one feminine
 - D) matched for level of control

17. Levenson (1973) suggested that locus of control was a multidimensional concept composed of:
 - A) externality, internality, and chance
 - *B) internality, powerful others, and chance
 - C) self-concept, self-esteem, and self-competence
 - D) esteem, depression, and preoccupation
18. Furnham (1987) administered seven scales of locus of control and found:
 - A) high alpha coefficients but low scale intercorrelations
 - B) a stable factor structure, high alpha coefficients, and five factors
 - C) factorial invariance and little item overlap among scales
 - *D) little item overlap, significant correlations, and low alphas
19. The Sexuality Scale was developed to:
 - A) assess whether couples are compatible or not
 - *B) measure what people think about their sexuality
 - C) identify sexual disturbances in psychiatric patients
 - D) assess such aspects as homosexuality and virginity
20. Is the sexual depression scale a measure of more global depression?
 - *A) The very limited evidence suggests no.
 - B) The extensive available evidence suggests no.
 - C) The very limited evidence suggests yes.
 - D) The extensive available evidence suggests yes.
21. The intimacy permissiveness scale
 - *A) is a good example of a Guttman scale
 - B) is one of the subscales of the Sexuality Scale
 - C) measures locus of control as applied to sexual behavior
 - D) has very low coefficients of reproducibility
22. Intelligence tests are to _____ as creativity tests are to _____.
 - A) divergent thinking; convergent thinking
 - B) internal locus of control; external locus of control
 - *C) convergent thinking; divergent thinking
 - D) factorial validity; tertiary validity
23. The second stage of the creative process:
 - A) preparation
 - B) illumination
 - C) verification
 - *D) incubation
24. Creative press involves:
 - A) the “aha” experience when solutions emerge
 - *B) both the inner psychological environment and the physical environment
 - C) creative writing, particularly poetry
 - D) making up stories as a way of measuring fantasy
25. Most of the available measures of creativity have:
 - A) high reliability and low validity
 - B) poor reliability and poor validity
 - C) excellent reliability and good validity
 - *D) unknown reliability and validity
26. The Torrance Test of Creative Thinking:
 - A) is a true-false inventory designed to self-assess creativity
 - *B) is actually a battery of verbal and figural subtests
 - C) measures to what degree a person possesses a creative personality
 - D) was originally developed for adults but is also used for children

Normal Positive Functioning

27. The subtests of the Torrance can be scored along the dimensions of:
- *A) fluency, flexibility, originality, and elaboration
 - B) divergent thinking and convergent thinking
 - C) preparation, incubation, illumination, and verification
 - D) identity, self-satisfaction, and behavior
28. In regards to the relationship between Torrance test scores and intelligence test scores:
- A) there is a pattern of very high correlations
 - B) the relationship is substantial but only in children
 - *C) there is a pattern of low correlations
 - D) there is a significant relationship but only in the elderly
29. The norms on the Torrance:
- A) were selected on the basis of U.S. census data
 - *B) seem to be samples of convenience, heterogeneous, and small
 - C) represent several thousand children from California
 - D) are quite extensive and comprised of large samples
30. Scores on the various subtests of the Torrance:
- A) are highly reliable
 - B) are easily obtained by hand-scoring
 - *C) intercorrelate substantially
 - D) are based on a scale of 100
31. In Guilford's model, the operation of _____ is most closely related to creativity.
- A) convergent thinking
 - B) memory
 - C) evaluation
 - *D) divergent production
32. "What would happen if people had three legs?" illustrates:
- *A) consequences test
 - B) making objects
 - C) transformations
 - D) convergent thinking
33. The Alternate Uses Test asks the subject to:
- A) define how common objects are used
 - *B) indicate how a common object might be used
 - C) solve problems that are made up of matchsticks
 - D) think of objects that can be used in different ways
34. Guilford's tests are typically characterized by:
- A) highly creative instructions
 - B) one correct answer per item
 - *C) very brief time periods
 - D) lack of a theoretical framework
35. The Alternate Uses Test presumably measures:
- A) the creative process
 - B) creative personality traits
 - C) the ability to transform common into creative
 - *D) flexibility of thinking
36. For the Alternate Uses Test a particularly crucial question concerns _____ reliability.
- *A) inter-rater
 - B) test-retest
 - C) alternate forms
 - D) internal consistency

37. The Alternate Uses Test is scored for:
- A) creativity
 - *B) fluency
 - C) flexibility
 - D) originality
38. In general, the reliability of Guilford's tests of divergent thinking is:
- A) very low
 - *B) marginal
 - C) about .70 as expected
 - D) very high
39. Guilford is particularly concerned about _____ validity.
- A) tertiary
 - B) discriminant
 - C) concurrent
 - *D) factorial
40. In regards to validity, tests like the Alternate Uses Test:
- A) have excellent validity
 - B) have good stability but low internal consistency
 - *C) leave much to be desired
 - D) have low stability and high internal consistency
41. The Domino Creativity scale on the ACL was developed by having:
- A) faculty members indicate which items are characteristic of creativity
 - *B) faculty members check ACL items to describe students differing in creativity
 - C) highly creative subjects describe themselves
 - D) experts in creativity indicate which items should be included
42. The items that were selected for the ACL creativity scale:
- *A) are consonant with both empirical findings and theoretical expectations
 - B) did not match the theoretical expectations of the author
 - C) are clearly in line with Guilford's structure of intellect model
 - D) all have substantial face validity
43. In the study of 800 high school students who had filled out a self-descriptive ACL, the Creativity scale:
- A) differentiated between creative males and creative females
 - B) showed significant differences between science and art students
 - *C) differentiated between creatives and controls regardless of field
 - D) seemed valid for male students but less so for female students
44. In regards to the reliability of the ACL Creativity scale:
- *A) internal stability seems high
 - B) test-retest over long time periods is low
 - C) has not been determined
 - D) lots of evidence to indicate it is highly reliable
45. On the Myers-Briggs Type Indicator, the higher one scores on _____ presumably the more creative.
- A) judging-perceiving
 - B) extraversion-introversion
 - *C) sensation-intuition
 - D) thinking-feeling

Normal Positive Functioning

46. The regression equation used to calculate the creativity index on the MBTI gives the greatest weight to _____ and the least weight to _____ .
- *A) SN; TF
 - B) JP; EJ
 - C) EI; TF
 - D) TF; SN
47. On the regression equation used to calculate the creativity index on the MBTI, two variables are given negative weights; they are:
- A) SN & TF
 - *B) EI & TF
 - C) SN & JP
 - D) JP & EI
48. On the regression equation used to calculate the creativity index on the MBTI, two variables are given negative weights. This means that the _____ one scores on these variables, the _____ .
- A) lower; less creative
 - B) higher; more valid is the regression equation
 - C) lower; lower one scores on all variables
 - *D) higher; less creative
49. The Chinese Tangrams
- *A) is equally applicable to children and adults, and does not require extensive verbal skills
 - B) requires extensive verbal skills such as those found in college students
 - C) does not seem to be particularly valid with children
 - D) requires the child to solve problems with Chinese characters
50. The directions on the Chinese Tangrams task:
- A) ask the client to reproduce a given completed figure
 - *B) present the test as a game
 - C) require the subject to solve combinatorial problems
 - D) present the task as a measure of intelligence
51. The responses on the Chinese Tangrams test are scored for:
- A) creativity, originality, and imagination
 - B) both convergent and divergent thinking
 - *C) fluency, flexibility, and originality
 - D) productivity and technical competence
52. For most tests of creativity _____ reliability seems more than adequate.
- A) test-retest
 - B) factorial
 - *C) interscorer
 - D) internal consistency
53. Intrascorer reliability refers to:
- A) do two different scorers agree on a specific protocol?
 - *B) Is the same scorer consistent over time?
 - C) Can the same scorer give identical scores to different protocols?
 - D) Do two different scorers achieve the same level of alpha reliability?
54. As a general statement, interrater reliability in psychological testing is _____ as compared to other fields like medicine.
- *A) higher
 - B) lower
 - C) equal
 - D) not as good

55. Basically there are two methods to assess originality:
- A) fluency and flexibility
 - B) adjective check lists and personality inventories
 - C) convergent and divergent thinking
 - *D) statistical infrequency and subjective judgment
56. Runco and Mraz (1992) argued that creativity is best measured “in toto.” Their results indicate:
- A) high convergent and discriminant validity
 - B) high discriminant validity but poor internal consistency
 - *C) high inter-rater reliability but poor discriminant validity
 - D) no relationship between ratings of creativity and of intelligence
57. Most measures of imagery are
- *A) self-report questionnaires
 - B) ratings by others
 - C) true-false inventories
 - D) forced choice questionnaires
58. In the Marks’ Vividness of Visual Imagery Questionnaire, the subject is asked:
- A) what it is they can imagine most and least vividly
 - *B) to picture a number of specific items and rate their vividness
 - C) to rate a number of pictures as to their vividness
 - D) a number of true-false questions about their imagery
59. The validity evidence for Marks’ Vividness of Visual Imagery Questionnaire indicates that:
- A) artists and other creative persons score higher
 - B) there are substantial correlations with MMPI scales
 - C) creative females score higher than creative males
 - *D) good visualizers are more accurate in their recall of colored photographs
60. Paivio’s Individual Differences Questionnaire attempts to measure:
- A) both creativity and originality
 - *B) both nonverbal imagery and verbal processes
 - C) the vividness of images in one’s “mind’s eye”
 - D) how vivid one’s dreams are
61. The inclusion or exclusion of items on Paivio’s Individual Differences Questionnaire was based on:
- *A) logical reasons
 - B) factor analyses
 - C) regression equations
 - D) criterion-keying
62. In regards to the internal consistency of the Individual Differences Questionnaire, most available coefficients are in the
- A) .30s
 - B) .60s
 - *C) .80s
 - D) .90s
63. The items on the Verbalizer-Visualizer Questionnaire were originally selected because they:
- A) showed substantial correlations with various criteria of imagery
 - *B) discriminated between individuals who used different eye movements
 - C) reflected Paivio’s dual coding theory
 - D) had substantial test-retest correlations

Normal Positive Functioning

64. A general concern about imagery questionnaires:
- *A) lack of consistent correlation between test scores and actual task performance
 - B) they depend on self-reports that may be dishonest
 - C) heavily affected by social desirability
 - D) they lack a theoretical framework
65. Items for the Smith and Houston (1992) competitiveness index were retained on the basis of:
- A) correlation with a scale of achievement motivation
 - B) factor loadings on one of three factors
 - *C) item-total correlations
 - D) low discriminant validity between athletes and controls
66. In the development of the Miller Hope Scale, items were deleted or retained on the basis of:
- A) authors' judgment
 - B) item-total correlations
 - C) high alpha coefficients
 - *D) expert judgment
67. A potential problem for the Miller Hope Scale is that scores on the scale
- A) have low internal consistency and stability
 - *B) correlate substantially with measures of psychological well being
 - C) correlate negatively with a hopelessness scale
 - D) correlate .69 with a one-item self-assessment of hope
68. There are two components on the Snyder Hope scale:
- A) satisfaction and anticipation of the future
 - *B) goal directed agency and pathways
 - C) hassles and uplifts
 - D) emotion and anxiety
69. The construct validity of the Snyder Hope scale is supported by:
- A) significant gender differences
 - *B) factor analytic studies
 - C) high internal consistency
 - D) the brevity of the scale
70. In answering the Hassles scale, the instructional time frame covers:
- A) yesterday
 - B) a typical week
 - *C) last month
 - D) up to now
71. The items in the Brief College Student Hassles Scale were originally:
- A) written by the authors
 - B) compiled from biographical sketches
 - C) obtained from Counseling Center clients
 - *D) written by Canadian college students
72. The UCLA Loneliness scale is characterized by:
- A) all positively worded items
 - *B) high coefficient alpha
 - C) strong factorial invariance
 - D) negative correlations with self ratings

73. The items for the scale of driving anger:
 - A) empirically discriminated drivers with good records and criminal records
 - B) were selected because they showed significant gender differences
 - *C) were based on interviews with college students and faculty
 - D) are all loaded on a factor of "road rage"
74. Originally, the items of the Death Anxiety scale were
 - A) selected from various anxiety scales
 - B) contributed by hospital chaplains
 - C) obtained from nursing home patients
 - *D) written by the author
75. The _____ translation method is used to assess the adequacy of translation of an instrument.
 - A) alternate
 - *B) back
 - C) convergent
 - D) direct

9 Special Children

1. The Education for all Handicapped Children Act specified
 - A) that screening and diagnosis be kept separate
 - *B) that linguistic differences should not interfere with the assessment of intelligence
 - C) all children be screened for neurological disorders
 - D) which tests of cognitive functioning were to be used
2. One of the problems of testing special children is that
 - A) tests do not mirror the school curriculum
 - B) there are few if any criterion-referenced tests
 - C) appropriate tests are not revised often enough
 - *D) diagnostic criteria are not well formulated
3. The textbook argues that as far as testing exceptional children:
 - *A) psychometrically valid instruments given by a trained professional can be quite useful
 - B) the majority of available tests are rather useless
 - C) impressionistic scoring is a far better procedure than standard psychometric approaches
 - D) there is clear-cut agreement as to what needs to be done
4. Although the area of _____ would seem to be of particular importance with special children, it is one that is relatively neglected.
 - A) cognitive
 - B) language
 - *C) personal-social
 - D) school readiness
5. The very procedure of testing a young child can be harmful and traumatic.
 - A) Most school psychologists would agree.
 - *B) There is little evidence to support this.
 - C) There is considerable evidence that this is so.
 - D) Various studies of multiple choice tests support this.
6. Infant intelligence tests are often assessed from the viewpoint of _____ validity.
 - *A) content
 - B) construct
 - C) criterion
 - D) face
7. The predictive validity of infant intelligence tests is low in part because:
 - A) infant intelligence is heavily influenced by verbal skills
 - *B) infant intelligence is exhibited primarily through motor behavior
 - C) such tests do not measure fine motor skills
 - D) most test items do not measure general maturation

8. The Gesell Developmental Schedules basically are a(n):
 - *A) timetable of expected behaviors
 - B) infant intelligence test
 - C) checklist to be filled out by parents
 - D) way of assessing social-emotional behavior
9. Most tests of preschool intellectual functioning use items based on the work of Gesell, even though
 - A) such items appear to have rather low validity
 - *B) Gesell did not believe that intelligence could be assessed
 - C) their test-retest reliability is low because of normal development
 - D) the Gesell Developmental Schedules are not strictly a test
10. The items on the Gesell Developmental Schedules are organized in terms of three “maturity zones”:
 - A) birth to one year, one year to 3 years, and 3 years and above
 - B) cognitive, motor, and social-emotional
 - *C) supine, sitting, and locomotion
 - D) mental scale, motor scale, and infant behavior record
11. The Gesell Developmental Schedules yields _____.
 - A) a ratio IQ
 - B) scores on seven dimensions
 - C) a mental IQ and a motor IQ
 - *D) five developmental quotients
12. The Infant Behavior Record is actually part of the:
 - *A) Bayley Scales of Infant Development
 - B) Gesell Developmental Schedules
 - C) Personality Inventory for Children
 - D) Vineland Adaptive Behavior Scale
13. The Infant Behavior Record is a(n)
 - *A) rating scale completed by the examiner at the end of testing
 - B) assessment of both gross and fine motor skills
 - C) checklist of items filled out by the parents
 - D) form of the Vineland Adaptive Behavior Scale
14. Visual tracking and auditory localization are part of which Bayley subscale?
 - *A) mental scale
 - B) motor scale
 - C) Infant Behavior Record
 - D) preschool affect scale
15. The administration of the Bayley Scales of Infant Development:
 - A) should be done by a pediatrician
 - B) can be carried out at home by parents
 - *C) requires a well trained examiner
 - D) is basically a structured interview of the parents
16. In administering the Bayley, the examiner
 - A) administers all items regardless of infant’s age
 - B) starts with the ceiling level and works backwards
 - *C) usually starts at 1 month below the infant’s CA
 - D) usually begins with the child’s mental age

17. Both the Mental Development Index and the Psychomotor Development Index of the Bayley have a mean of:
 - A) 50
 - *B) 100
 - C) 500
 - D) 16
18. According to its author, the Bayley should be used to:
 - *A) assess current developmental status
 - B) predict future academic achievement
 - C) identify neurological problems that may develop
 - D) assess general intelligence
19. There is not much validity information available on the Infant Behavior Record because:
 - A) the IBR is too new to have been evaluated
 - *B) the behavior measured is specific to the test situation
 - C) it is quite difficult to assess infants validly
 - D) the IBR measures global personality variables
20. The norms on the Bayley scales:
 - A) include infants, children, and adolescents
 - *B) are presented in half month and one month intervals
 - C) were based on samples of institutionalized children
 - D) are based on a large sample, ages 2 months to 12 years
21. The measurement of social-emotional behavior in infants and young children has received less emphasis because:
 - A) of Public Law 94-142 and other legislation
 - B) parents are more interested in their child's intelligence
 - C) tests like the Vineland are not well known
 - *D) there was a greater need to assess cognitive abilities
22. Other than the Bayley, most scales that measure the social-emotional aspects of infants:
 - A) can be administered by parents
 - *B) are weak in their construct validity
 - C) are true-false inventories
 - D) have very low reliability
23. The Personality Inventory for Children:
 - *A) is filled out by an adult informant such as mother
 - B) is filled out directly by the child with help from an adult
 - C) yields a set of scales parallel to the MMPI
 - D) can be administered in about 10 minutes or less
24. An interesting aspect of the Personality Inventory for Children is that:
 - A) test-retest reliability is rather low
 - B) there is at present no validity data
 - *C) the test can be administered entirely or in portions
 - D) the child responds to each item by pointing to "happy" faces
25. In regards to the norms of the Personality Inventory for Children, they are:
 - *A) based on a large sample but basically from one city
 - B) carefully chosen and representative of census data
 - C) available for a large number of varied subgroups
 - D) made up of small, atypical samples

26. The Minnesota Preschool Affect Rating Scale:
- *A) assesses videotaped behavior of a child with parent
 - B) is used to assess a group of children on the playground
 - C) is specifically designed for epileptic children
 - D) yields an overall mental score and a motor score
27. Concordance rate:
- A) how close MA and CA are to each other
 - *B) degree of agreement between raters
 - C) the degree of variance in one variable explained by a test
 - D) whether twins are fraternal or identical
28. The book states that an examiner “should be friendly but calm” when testing younger children because:
- A) children respond to smiles and giggles
 - B) the parents will be more cooperative
 - *C) too much enthusiasm can scare or overwhelm a child
 - D) calm examiners establish better rapport
29. In regards to testing of exceptional children, the training that psychologists receive is:
- A) generally excellent, especially if they are clinical psychologists
 - B) usually thorough, with special emphasis in mental retardation
 - C) excellent, especially if they are in an APA approved program
 - *D) very limited and insufficient
30. If a young child is to be tested, the test(s) selected should
- A) be the most current, even if the examiner is not familiar with them
 - *B) be pertinent to the referral question
 - C) have a wide band-width
 - D) be selected because of their high reliability
31. A 28 year old, victim of an automobile accident, now has an IQ of 42, but he would not be labelled as mentally retarded because:
- *A) the IQ deficit did not show up as a child
 - B) only children can be thus classified
 - C) we don't know what his pre-accident IQ was
 - D) there is brain damage present, not retardation
32. Adaptive behavior refers to:
- A) what is measured by an intelligence test
 - B) the behavior of a child who is mentally retarded
 - *C) effectiveness of an individual in coping
 - D) the responses made by a child to an adaptive test
33. The Survey form and the expanded form of the Vineland differ in that:
- A) the expanded form is identical to the classroom edition
 - B) the two forms yield different scales
 - C) the Survey form only measures daily living skills
 - *D) the expanded form is a longer version of the Survey form
34. Both the Survey form and the expanded form of the Vineland are completed by a(n) :
- A) parent
 - B) teacher
 - *C) examiner
 - D) client

35. Essentially, the Vineland
- *A) requires a semi-structured interview of a parent
 - B) consists of true-false items read by the client
 - C) is much like a Stanford-Binet with a series of subtests
 - D) is an observation checklist to be filled out by a teacher
36. In the revised Vineland raw scores are changed to standard scores with a mean of _____. .
- A) 0
 - B) 15
 - C) 50
 - *D) 100
37. The Vineland assesses all of these except:
- A) motor skills
 - *B) neuropsychological status
 - C) communication
 - D) daily living skills
38. The various forms of the Vineland
- A) are identical to each other
 - B) basically differ in their difficulty level
 - *C) all yield the same scales and subscales
 - D) differ as to whether they measure communication skills
39. The Vineland is administered
- *A) like an interview, to a parent or other informant
 - B) like a personality test, with a specified item sequence
 - C) by asking the child directly a series of questions
 - D) like a WISC, with subtests and time limits
40. Because the Vineland consists of simply worded items, an examiner needs
- A) only the ability to read the items
 - *B) to be well trained, as in administering a Stanford-Binet
 - C) almost no training at all
 - D) only to establish rapport
41. The Vineland yields scores
- *A) similar to the IQ obtained on an intelligence test
 - B) similar to the MMPI clinical scales
 - C) that reflect the degree of pathology in the client
 - D) that measure the degree of receptive vocabulary
42. For the Vineland, the lowest reliability coefficients are for:
- A) split-half reliability
 - *B) inter-rater reliability
 - C) test-retest reliability
 - D) inter-scorer reliability
43. The evidence suggests that in regard to cognitive functioning the Vineland
- A) is basically a test of intelligence
 - B) is an excellent measure of cognitive functioning
 - *C) measures something different than, yet related to cognitive functioning
 - D) measures skills that are not related to cognitive functioning
44. From a validity point of view, on the Vineland:
- A) the Survey form seems to be weak
 - B) all three forms have marginal validity
 - C) all three forms have excellent validity
 - *D) the classroom edition seems to be weak

45. The manual of the Vineland contains information so it can be used with the
A) Stanford-Binet
B) WISC
C) Peabody Picture Vocabulary
*D) K-ABC
46. The norms for the revised Vineland cover the ages:
A) 9 to 18 years
B) 2 to 6 years
*C) birth to 19 years
D) birth to 6 years
47. The norms for the revised Vineland:
*A) cover both normal and special children
B) cover only mentally retarded adults
C) are clearly based on samples of convenience
D) are identical to the Stanford-Binet norms
48. Of all the various categories of special children, the most filled with controversy as to definition and scope:
A) behavioral-emotional disorders
*B) learning disabilities
C) motor impairments
D) hearing impairments
49. A wide range of conditions involving muscle degeneration and weakness:
A) spina bifida
*B) muscular dystrophy
C) cerebral palsy
D) autism
50. The incidence of mental retardation among those who have cerebral palsy is about:
A) 10%
*B) 50%
C) 80%
D) 95%
51. In assessing children with cerebral palsy, the examiner needs to be sure that
A) at least one parent is present during testing
*B) the testing situation is as objective as possible
C) the child is also not mentally retarded
D) the test used emphasizes motor rather than verbal abilities
52. An intelligence test likely to be used without modification for a child with motor impairments:
A) K-ABC
B) Boehm test
C) Hiskey-Nebraska
*D) Peabody Picture Vocabulary Test
53. In testing children with motor impairments, some examiners prefer the Stanford-Binet over the WISC-R because:
*A) of its modifiability and higher proportion of verbal items
B) it yields a profile of subscores
C) it is much more valid than the WISC-R
D) special scales have been developed on it to measure motor skills

54. This test consists of line drawings to which the child points to match a drawing to a word.
- A) Illinois Tests of Psycholinguistic Abilities
 - B) Hiskey-Nebraska tests
 - *C) Peabody Picture Vocabulary Test
 - D) Vineland
55. A “special” aspect of the Peabody Picture Vocabulary Test:
- A) has a brief time limit
 - B) all the items are administered to every child
 - C) requires a well trained examiner
 - *D) has two forms
56. The items on the Peabody were originally selected
- A) from an analysis of popular children’s books
 - *B) by the author to represent common words
 - C) by asking parents what words their children knew
 - D) by tape recording actual conversation with children
57. Raw scores on the peabody are converted to:
- A) a ratio of correct to incorrect responses
 - B) T scores with a mean of 50 and SD of 10
 - C) “decision” scores to determine whether a child is a false negative
 - *D) standard scores with mean of 100 and SD of 15
58. As far as the items on the Peabody, research shows that:
- *A) potentially biased items have been eliminated
 - B) many of the items show racial or cultural bias
 - C) one needs to be very cautious when testing minority children
 - D) at least 20% of the items show a gender bias
59. On the Peabody, the standard error of measurement:
- A) is a reflection of the high reliability of the test
 - *B) is about twice the size of those on standard tests of intelligence
 - C) is substantially smaller than expected suggesting low validity
 - D) is a good measure of the alternate form reliability
60. As a very general conclusion, we can say that the Peabody correlates _____ with tests of intelligence.
- A) .05 to .25
 - B) in the low .20s
 - *C) .70s and low .80s
 - D) about .90
61. The evidence indicates that the Peabody:
- A) is a better alternative than the Vineland
 - B) tends to yield higher average scores than the WISC-R
 - C) has poor concurrent validity
 - *D) should not be substituted for a more comprehensive measure
62. in regards to adult clients, the Peabody:
- A) is very valid but only with the mentally retarded
 - B) is quite useful with both normal and retarded adults
 - *C) the results are mixed, and caution should be used
 - D) is simply not appropriate with clients older than 15
63. Norms for the revised Peabody:
- *A) are quite extensive and based on census data
 - B) are based only on samples of children
 - C) seem to be samples of convenience, small and atypical
 - D) are not available at this time

64. The study by Hayes and Martin (1986) using the Peabody versus the Stanford-Binet to identify gifted children concluded that:
- A) the Peabody was an excellent measure
 - *B) using the Peabody was questionable
 - C) the Stanford-Binet was much more useful
 - D) both measures could be used
65. The literature suggests that the Peabody
- A) yields a reliable and valid verbal IQ
 - B) can substitute for the Stanford-Binet when time is of the essence
 - *C) is most useful as a screening instrument
 - D) has very low reliability and hence limited validity
66. Speaking is considered _____ language, while reading is considered _____ language.
- A) receptive; expressive
 - B) both are receptive
 - C) both are expressive
 - *D) expressive; receptive
67. Early identification of language delay in children is very important because language delay
- A) is a good indicator of the presence of cerebral palsy
 - *B) can result in academic and social difficulties
 - C) if not corrected results in autism
 - D) results in hearing impairments
68. The Boehm Test of Basic Concepts assesses a child's
- *A) mastery of concepts like left and right
 - B) basic vocabulary like "shirt" and "pie"
 - C) ability to follow directions correctly
 - D) cognitive deficits
69. There is a version of the Boehm available for:
- *A) blind children
 - B) Chinese speaking children
 - C) children with spina bifida
 - D) normal adults
70. The two alternate forms of the Boehm use _____ concepts and _____ illustrations.
- A) different; same
 - *B) same; different
 - C) different; different
 - D) same; same
71. Which of these accurately characterizes the Boehm?
- A) its black and white drawings are very ambiguous
 - B) the examiner needs to be extremely well trained
 - C) the questions are answered by the parent of the child
 - *D) can be administered to a small group of children
72. A problem with the scoring of the Boehm:
- A) requires expertise in basic statistics, like z scores
 - B) can only be machine scored by the publisher
 - *C) time consuming if many protocols are to be scored
 - D) scoring is quite complicated with complex directions

73. The reliability of the Boehm is low for second grade and older children. The reason for this:
- A) the test is very difficult for second graders
 - B) the test is very short and hence unreliable
 - C) the distribution of scores has a high positive skew
 - *D) a ceiling effect
74. Test items on the Boehm were chosen on the basis of:
- *A) frequency of use in school
 - B) substantial criterion-related validity
 - C) high test-retest reliability
 - D) expert judgment
75. The norms for the Boehm:
- A) are only available for small, atypical samples
 - *B) are fairly substantial and reflect census data
 - C) range from age 1 to age 21
 - D) are excellent for such special children as the retarded
76. The categories of hard of hearing and deaf differ as to whether
- A) the impairment resulted before or after language development
 - *B) hearing level can be enhanced
 - C) the person is congenitally or adventitiously deaf
 - D) the person had a viral infection
77. A person who is adventitiously deaf
- *A) became deaf later in life
 - B) was born deaf
 - C) can be helped with a hearing aid
 - D) is deaf because of trauma
78. Of children enrolled in special ed programs
- A) most are adventitiously deaf
 - *B) 2 out of 3 have a serious hearing loss
 - C) 2 out of 3 are postlingually deaf
 - D) 1 out of 10 have a profound or severe hearing loss
79. Rapport is especially important in testing hearing impaired children because
- A) they can only understand through lip reading
 - *B) they are often socially withdrawn and shy
 - C) rapport enhances the reliability of a test
 - D) verbal tests are particularly sensitive to rapport
80. McQuaid and Aloviseti (1981) in their survey of services for hearing impaired found that the _____ was commonly used.
- A) verbal scale of the WAIS-R
 - B) vocabulary items from the Stanford-Binet IV
 - C) Picture completion subtest from the WISC
 - *D) Performance scale of the WISC-R
81. Originally, the Hiskey-Nebraska Tests of Learning Aptitude were standardized on:
- A) adventitiously deaf children
 - *B) hearing children
 - C) hearing-impaired children
 - D) carefully defined institutional samples

82. The Hiskey-Nebraska is applicable to:
- A) both children and adults
 - B) children younger than 5
 - *C) ages $2\frac{1}{2}$ to $17\frac{1}{2}$ years
 - D) any age for hearing impaired
83. The deaf children that were tested with the Hiskey-Nebraska for its subsequent standardization:
- A) were carefully chosen on the basis of census data
 - *B) were likely samples of convenience
 - C) contained equal numbers of pre- and post-lingual children
 - D) were well defined clinical groups
84. On the Hiskey-Nebraska
- A) all items are presented to all children
 - *B) which items are presented is a function of the child's age
 - C) which items are presented reflect the child's vocabulary score
 - D) all items are presented in pantomime
85. We would expect the administration of the Hiskey-Nebraska to last about:
- *A) 1 hour
 - B) 2 hours
 - C) 10 to 15 minutes
 - D) over 4 hours
86. Median scores are used on the Hiskey-Nebraska subtests because:
- A) the reliability of the subtests is marginal
 - *B) the author's belief that deaf children score poorly on initial items
 - C) the median should be used with skewed distributions
 - D) deaf children often score below the median
87. As a general comment, we can characterize the Hiskey-Nebraska as an:
- A) alternate form of the WISC
 - B) adaptation of the Wechsler series
 - C) comparable form of the Peabody
 - *D) adaptation of the Stanford-Binet
88. Test experts generally feel that the Hiskey-Nebraska
- A) is an excellent instrument for hearing-impaired children
 - *B) is psychometrically inadequate
 - C) should be used with adults but not with children
 - D) has excellent reliability as well as validity
89. The Snellen chart measures
- *A) visual acuity
 - B) cognitive impairment
 - C) degree of hearing loss
 - D) motor incoordination
90. Most of the tests used with the visually handicapped have:
- A) well chosen representative norms
 - B) excellent reliability and validity
 - *C) been standardized on normal samples
 - D) small standard errors of measurement
91. The most frequently used cognitive test with visually impaired clients:
- *A) Wechsler scales
 - B) Hayes-Binet
 - C) Grassi Basic Cognition Evaluation
 - D) Hiskey-Nebraska

92. A major criticism of the procedures used for the identification of the gifted or talented:
- A) based on teacher's ratings
 - B) lack reliability and validity
 - C) are biased against female children
 - *D) focus on convergent rather than divergent thinking
93. Checklists of behaviors related to giftedness
- A) are very useful as screening instruments
 - B) typically have high interscorer reliability
 - *C) often lack reliability and validity
 - D) are not readily available
94. In regards to the readability of self-report tests:
- A) the available indices of readability have poor reliability
 - B) the readability level is typically appropriate
 - *C) it is quite high for many measures
 - D) it is higher on tests like the Peabody
95. Testing the limits can be used as a strategy to
- *A) assess learning potential
 - B) measure visual acuity
 - C) determine whether a child is autistic
 - D) assess test-retest reliability
96. Test-retest stability of preschool measures is typically
- *A) low because behaviors are unstable at those ages
 - B) low because most measures have low reliability
 - C) high because most measures have adequate validity
 - D) high because young children are consistent
97. The two major approaches to neuropsychological assessment of children are to either use a standardized battery like the Luria-Nebraska or to use
- A) medical procedures like brain scans
 - *B) a combination of traditional tests like the K-ABC plus others
 - C) the Reitan battery
 - D) multivariate instruments like the MMPI
98. The tasks used in the Luria-Nebraska Children's battery are
- A) identical to the ones in the adult battery
 - B) downward extensions of adult items
 - *C) modeled on an adult version
 - D) totally unrelated to the adult version
99. The initial assignment of items to a specific subscale of the Luria-Nebraska Children's battery was done on the basis of:
- A) consensus among a panel of experts
 - B) factor analysis
 - C) content analysis
 - *D) author's clinical judgment
100. Each of the 11 scales in the Luria-Nebraska Children's battery is:
- A) a pure measure of a specific brain function
 - *B) a multifactorial measure of a set of related skills
 - C) comparable to a measure of "g" in intelligence
 - D) a measure of a specific neurological syndrome

101. Items on the Luria-Nebraska Children's battery are scored to reflect:
 - A) how far above the mean the child's performance is
 - B) the degree of personality disturbance
 - *C) how far below the mean the response is
 - D) how much agreement the child expresses
102. Q-factor analysis involves:
 - A) factor analysis of primary factors
 - *B) identifying subgroups of subjects with similar test profiles
 - C) matching a factor structure to a hypothesized theory
 - D) predicting which type of child has a learning disability
103. The construct validity of the Luria-Nebraska Children's battery is:
 - *A) judged to be poor
 - B) excellent, because of Luria's theory
 - C) well supported by studies both in Russia and the US
 - D) yet to be determined
104. Drawing techniques were initially developed as measures of:
 - A) personality
 - B) social skills
 - *C) cognitive functioning
 - D) fantasy
105. In regards to drawing techniques, we can say as a very general conclusion that:
 - *A) such tests have little value in assessing personality
 - B) drawing skills do not affect performance
 - C) they have poor reliability but good validity
 - D) in the hands of an expert they are as valid as the K-ABC
106. The Developmental Test of Visual-Motor Integration measures
 - A) a child's proficiency in communication skills
 - B) the presence and location of minor brain damage
 - C) learning ability of deaf children
 - *D) problems in visual-motor or visual-spatial abilities
107. In general, the predictive validity of tests like the VMI is better with:
 - A) older children than younger children
 - B) adults than children
 - *C) younger children than older children
 - D) males than females
108. One of the major limitations of behavior rating scales:
 - *A) low interrater agreement
 - B) require substantial training
 - C) do not provide information on infrequent behaviors
 - D) are less reliable than unstructured interviews
109. The various forms of the Child Behavior Checklist differ
 - *A) as to who the rater is and the age of the child
 - B) in the type of variables that are scaled
 - C) only as to which informant is the scale intended for
 - D) as to whether they measure overcontrol or undercontrol
110. In regards to the validity of the Child Behavior Checklist:
 - A) generally poor and not supportive of the theoretical framework
 - B) the manual does not give sufficient detail
 - C) literature results are inconsistent
 - *D) generally good and supportive of the construct validity

Special Children

111. Both the Conners rating scales and the Child Behavior Checklist:
- A) have unknown validities
 - B) were developed for research purposes only
 - *C) provide separate scales for parents and for teachers
 - D) are basically measures of hyperactivity
112. On the Conners' Parent and Teacher forms:
- A) the items are identical
 - *B) the areas covered are different
 - C) the factor structure is the same
 - D) the response options are different
113. A criticism of the Conners scales:
- *A) the response choices are ambiguous and ill defined
 - B) the true-false format does not yield sufficient variability
 - C) they are strongly affected by social desirability
 - D) alpha coefficients are low
114. A somewhat unusual feature of the McCarthy Scales of Children's Abilities:
- A) the 18 subtests maximize the child's anxiety
 - B) it does not require a well trained examiner
 - *C) test materials and procedure are designed to maximize a child's performance
 - D) contains significant gender differences
115. There seem to be three major factors in the McCarthy Scales:
- A) verbal, quantitative, and perceptual-performance
 - B) block building, word knowledge, and leg coordination
 - C) antisocial, psychosomatic, and anxious-inhibited
 - *D) general cognitive, memory, and motor
116. The norms on the McCarthy Scales were based on:
- A) normal children, primarily samples of convenience
 - *B) normal children selected according to census characteristics
 - C) institutionalized children, especially mentally retarded
 - D) children with various deficiencies

10 Older Persons

1. The concept of functional age refers to:
 - A) the age at which a child learns a task like walking
 - B) how a person functions in a stressful situation
 - *C) a person's ability to be involved in various activities
 - D) the rate at which intellectual functions deteriorate
2. Many measures used with older persons were originally developed using younger samples:
 - *A) but most show equivalence in reliability and validity
 - B) with equivalence in factor structure but not reliability
 - C) who were then retested when they were older
 - D) and were therefore not concerned with social desirability
3. Face validity in testing older persons is important because:
 - *A) the client needs to feel that the test is appropriate and useful
 - B) older persons are, as a group, highly suspicious
 - C) face validity is equivalent to tertiary validity
 - D) face validity lessens fatigue and other problems
4. Which of these is NOT a procedure to be implemented when testing older persons?
 - A) give practice items
 - B) make sure the testing room is free of distractions
 - C) give lots of encouragement
 - *D) shorten the time limits so the client doesn't get tired
5. Performance assessment differs from self-assessment in that performance assessment
 - A) involves the performance scale of a test like the WAIS
 - B) is psychometrically superior to self-assessment
 - C) is more acceptable to the participants
 - *D) involves observing behavior directly
6. Myers et al. (1993) compared performance assessment and self assessment in a sample of elderly adults and found that:
 - A) most subjects preferred to carry out tasks than to answer questionnaires
 - *B) only about $\frac{1}{2}$ of the subjects attempted the performance tasks
 - C) the performance method was clearly superior to the self-assessment
 - D) both methods were unreliable
7. MMPI results with older persons indicate that:
 - A) mean scores on all scales increase by about one SD
 - B) mean scores on all scales decrease in varying amounts
 - *C) mean scores increase on some scales and decrease on others
 - D) mean scores on all scales essentially stay the same

8. In applying a test like the WAIS to older persons we need caution in that:
 - *A) available norms may not be appropriate
 - B) their reliability and validity is poor
 - C) equivalence of factor structure may not be there
 - D) these tests are not applicable to older persons
9. In the Kogan and Wallach (1961) study comparing attitudes and values of older persons with those of college students through use of a semantic differential, they found that
 - A) older subjects had great difficulty completing the task
 - *B) a factor analysis yielded a strong evaluative factor
 - C) college students were better educated and more intelligent
 - D) older subjects rated all concepts in a less favorable way
10. Guttman scales have been used with older persons especially in the assessment of:
 - A) sexual interest and marital satisfaction
 - B) attitudes towards retirement and death
 - *C) difficulties in performing activities of daily living
 - D) how older persons cope with declining health
11. In the area of attitudes towards older persons, the _____ has been used a number of studies.
 - *A) semantic differential
 - B) life satisfaction scale
 - C) adjective check list
 - D) brief mental status exam
12. Tuckman and Lorge (1952) developed a scale to measure attitudes of college students towards old age. They used a yes-no response format over a percentage format because:
 - A) it yielded a higher alpha coefficient
 - *B) the yes-no format took less time and simplified the instructions
 - C) older subjects preferred the yes-no format
 - D) the yes-no format yielded more valid results
13. The Tuckman and Lorge (1952) scale of attitudes towards old age
 - *A) lacks construct validity
 - B) lacks face validity
 - C) has been used in few studies
 - D) is now outdated
14. In the Axelrod and Eisdorfer (1961) study, different subgroups of students responded to different age groups (e.g., 35 vs 75) in filling out an attitude scale. The results showed:
 - A) few of the scale items showed monotonic increases
 - B) no significant differences
 - C) very poor interscorer reliability
 - *D) more negative stereotypes with older age groups
15. The Lasher and Faulkender (1993) anxiety about aging scale was developed using a theoretical model with two dimensions:
 - A) depression and anxiety
 - B) physical and psychological
 - C) fear of aging and search for meaning
 - *D) fears and dimensions of anxiety
16. An unusual aspect of the Lasher and Faulkender (1993) anxiety about aging scale:
 - A) graduate students rather than experts were used
 - *B) the initial pool of items was not reduced
 - C) there were no significant gender differences
 - D) the results of the factor analysis supported the two dimensions

17. Two basic approaches are used to measure the psychological well being of older people. One is to look at _____ and the other is _____.
 - A) the semantic differential; Neugarten life satisfaction scales
 - B) marital satisfaction; personal adjustment
 - *C) overt behavior; the person's own evaluation
 - D) attitudes towards aging; physical health
18. Initially, Neugarten and colleagues developed a life satisfaction set of ratings based on:
 - *A) interviews
 - B) MMPI profiles
 - C) expert judgment
 - D) self reports
19. Neugarten et al. (1961) defined psychological well being as composed of five aspects such as zest and positive self-concept. These aspects were
 - A) later shown to be relatively independent of each other
 - *B) intercorrelated substantially when rated through interviews
 - C) eventually operationalized into a set of five Life Satisfaction scales
 - D) the basis for developing the Philadelphia Morale Scale
20. The Life Satisfaction Index A and Index B:
 - A) measure morale and well-being respectively
 - B) are simply two versions of the same scale
 - *C) differ in the type of response required
 - D) were developed in rather different ways
21. In a national survey, Doyler and Forehand (1984) found that lowered Life satisfaction reflected
 - A) chronological age as the major variable
 - B) whether a person was married or not
 - C) the coping strategies used by the individual
 - *D) poor health, loneliness, and money problems
22. The Carstensen and Cone (1983) study of life satisfaction and morale in a sample of older persons suggests that such measures
 - *A) may be affected by social desirability
 - B) have good reliability but poor validity
 - C) are assessing independent variables
 - D) show significant gender differences
23. The initial pool of items for the Marital Satisfaction Questionnaire for Older Persons came from
 - A) interviews of happily married couples
 - *B) a variety of sources
 - C) observations of divorce proceedings
 - D) expert judgment
24. The response format used on the Marital Satisfaction Questionnaire for Older Persons uses
 - A) a standard Likert scale
 - B) a true-false format, easier for older people
 - *C) a six point scale with no central response
 - D) forced choice adjectives
25. Items for the final version of the Marital Satisfaction Questionnaire for Older Persons were retained on the basis of:
 - *A) distribution of responses and internal consistency
 - B) discriminant validity between married and divorced groups
 - C) judgment by experts in marital counseling
 - D) test-retest item correlations above .70

26. As part of the development of the Marital Satisfaction Questionnaire for Older Persons, it was correlated with the:
- *A) Locke-Wallace Marital Adjustment Test
 - B) Life Satisfaction Index B
 - C) Philadelphia Geriatric Center Morale Scale
 - D) Death Images Scale
27. In general, the Marital Satisfaction Questionnaire for Older Persons seems to have:
- A) no gender differences
 - *B) good construct validity
 - C) a factor structure of five major factors
 - D) poor internal consistency
28. Scales of morale can be generally criticized as:
- A) too short and therefore not reliable
 - *B) lacking conceptual clarity and adequate reliability
 - C) heavily influenced by Freudian notions of conscience
 - D) not easily translated from one language to another
29. The 22 items for the Philadelphia Geriatric Center Morale Scale were retained from a larger pool of items
- A) on the basis of construct validity
 - B) since they showed significant loadings on a factor of morale
 - *C) because they correlated with rankings of two professionals
 - D) based on a content analysis
30. The Philadelphia Morale Scale is purposely short
- *A) so as not to fatigue the respondent
 - B) because few items survived the statistical analyses
 - C) as a way of enhancing reliability
 - D) because the initial pool of items was quite small
31. In regards to the factor structure of the Philadelphia Morale Scale:
- A) a factor of dissatisfaction seems to be primary
 - *B) different studies and versions of the scale yield different factors
 - C) a five factor solution seems the best
 - D) the results seem to be a function of age
32. In regards to the reliability of the Philadelphia Morale Scale, we can conclude that:
- A) it is a reliable instrument with solid split-half reliability
 - B) the computed Cronbach alphas are rather low
 - *C) it is on the low side, in part because the scale is brief
 - D) reliability is problematic because of poor construct validity
33. In regards to the validity of the Philadelphia Morale scale, we can question whether
- A) the use of Q sorts was appropriate
 - B) the scale correlates with other measures of morale
 - C) there is invariance of factor structure
 - *D) the scale measures morale
34. About half of the items for the Geriatric Scale of Recent Life Events
- A) were taken from the MMPI depression scale
 - B) show substantial correlations with measures of morale
 - C) load on a factor of depression
 - *D) came from the Holmes and Rahe scale

35. The items on the Geriatric Scale of Recent Life Events
 - A) are true-false depression type items
 - B) require the respondent to indicate how satisfied they are
 - *C) represent everyday stresses or major life crises
 - D) involve logical memory and visual reproduction
36. The authors of the Geriatric Scale of Recent Life Events report substantial correlations between the stress weights in their study and those reported by Holmes and Rahe. They label this reliability; assuming they are correct what type of reliability might this be:
 - A) test-retest
 - *B) alternate forms
 - C) split-half
 - D) internal consistency
37. Marshall (1982) reviewed 32 instruments having to do with death and dying, and found the largest category to be measures of:
 - *A) death anxiety
 - B) the experience of death
 - C) death images
 - D) attitudes towards dying
38. The death images scale developed by Back (1971) uses as items:
 - A) vignettes
 - B) true-false
 - C) multiple choice
 - *D) metaphors
39. The most commonly used measure of memory functioning is the:
 - A) Mattis Dementia Rating Scale
 - *B) Wechsler
 - C) Halstead-Reitan
 - D) Social Readjustment Rating Scale
40. There are two basic questions relevant to the validity of measures of neuropsychological assessment of older persons. One is whether the test scores differentiate older persons with brain-damage from older persons with functional disorders. The other is:
 - *A) brain-damaged older persons vs. non-brain damaged older persons
 - B) brain-damaged older persons vs. middle-aged brain-damaged
 - C) well adjusted older persons vs. brain-damaged older persons
 - D) older persons with and without Alzheimer's
41. Dementia refers to:
 - A) a disease due to stroke
 - B) the condition also known as Alzheimer's
 - *C) progressive decline in intellectual functions
 - D) a condition characterized by depression and hallucinations
42. In general questionnaires that assess subjective memory failures in older adults have
 - *A) poor test-retest reliability
 - B) excellent internal consistency
 - C) been around since the days of William James
 - D) adequate reliability
43. The relationship between subjective reports of memory difficulty and actual performance on verbal memory tests is
 - A) substantial in all older persons
 - *B) a function of diagnostic status
 - C) positive, but only for Alzheimer's patients
 - D) positive, but only in depressed patients

44. Herrmann (1982) reviewed 14 memory questionnaires and concluded that _____, while your text concludes that _____.
- *A) they were reliable; reliability data was either absent or marginal
 - B) they were not reliable; they were reliable
 - C) reliability data was absent; their reliability is an open issue
 - D) their reliability was poor; their reliability was not proven
45. One of the limitations of memory tests is that they:
- A) use everyday tasks rather than laboratory procedures
 - B) are heavily influenced by theoretical models
 - *C) bear little resemblance to the tasks of everyday life
 - D) are too brief and hence unreliable
46. The Wechsler Memory Scale:
- *A) contains subtests that are combined into two summary scores
 - B) yields a profile of seven components of memory
 - C) is essentially a version of the WAIS for older persons
 - D) yields both an orientation score and a logical memory score
47. When the Wechsler Memory Scale was revised, the new version
- A) was essentially an alternate form for the original
 - B) seemed to be less valid than the original
 - *C) was a much improved and different scale
 - D) simply reflected some minor cosmetic changes
48. The Memory Assessment Clinics Self-Rating Scale differs from the Wechsler Memory Scale in that:
- A) the first is much longer than the second
 - B) the first yields one global memory index
 - *C) the first measures complaints, the second performance
 - D) the first is not affected by social desirability as is the second
49. The most common functional psychiatric disorder among older persons:
- A) anxiety about death
 - B) Alzheimer's
 - C) senile dementia
 - *D) depression
50. A potential confounding aspect of depression scales when used with older persons:
- A) older persons experience more sleeping difficulties
 - *B) items that reflect somatic symptoms
 - C) their depression may be a realistic response to old age
 - D) older persons are more likely to give negative responses

11 Testing in a Cross-Cultural Context

1. Most concerns about test bias are related to tests of:
 - A) attitudes
 - B) aptitudes
 - C) personality
 - *D) intelligence
2. On a test like the WAIS or other intelligence test, we would expect the average performance of blacks to be about _____ points below the average of whites.
 - A) 3
 - B) 10
 - *C) 15
 - D) 25
3. Bias can be potentially involved at three different stages. Which one is not one of these:
 - A) the test itself as a measuring instrument
 - *B) the computing of reliability and validity indices
 - C) the decisions made on the basis of test scores
 - D) in such aspects as racial discrimination
4. One of the reasons that minority children may do less well on tests of cognitive abilities may involve “irrelevant difficulty” on the test, such as:
 - A) marginal test-retest reliability
 - *B) an answer sheet that is confusing
 - C) poor norms
 - D) lack of construct validity
5. The cultural test bias hypothesis:
 - *A) group differences reflect artifacts of the tests
 - B) differences in performance are due to differences in culture
 - C) the bias resides in cultural aspects like poverty
 - D) we need to test the entire culture to assess bias
6. To the argument that tests are biased and should not be used with blacks, experts responded that:
 - *A) decisions still need to be made
 - B) the argument did not make any sense
 - C) tests are not biased
 - D) the evidence supports such an argument
7. To the argument that cognitive tests mirror middle class values, experts responded that:
 - A) minorities needed to be better represented in the norms
 - *B) there was no evidence that cognitive tests are biased
 - C) middle class values are indeed related to achievement
 - D) the argument was wrong and biased

8. To the argument that most examiners are white and this pushes down minority children's scores, we can respond that:
 - *A) the literature does not support such an argument
 - B) the same applies to white children tested by a black examiner
 - C) this simply mirrors a black child's inability to cope
 - D) the literature does support such an argument
9. Lam (1993) indicates that test developers make five assumptions such as children, whatever their race, can follow test directions. Such an issue:
 - A) has been neglected until now
 - B) proves that test developers are biased
 - C) is actually part of construct validity
 - *D) is incorporated in the test Standards
10. Matluck and Mace (1973) made a number of recommendations regarding tests to be used with Mexican-American children. In regards to drawings:
 - *A) sketchy line drawings should not be used
 - B) sketchy line drawings are excellent for young children
 - C) actual objects or photographs are always better
 - D) Mexican-American children are better at drawing
11. When bias is defined in terms of validity:
 - A) internal criteria involves item difficulty
 - *B) external criteria involves predictive validity
 - C) external criteria involves construct validity
 - D) internal criteria involves predictive validity
12. From an internal point of view, a test is unbiased if it:
 - A) loads on g
 - B) has equivalent regression equations
 - C) shows significant mean differences
 - *D) behaves the same way for different groups
13. Differential validity means that:
 - A) a test is free of bias
 - *B) a test works one way in a sample, but differently in another sample
 - C) both internal and external criteria are equally valid
 - D) one type of validity is high but another is low
14. In regards to differential validity, we can conclude that
 - *A) it does not seem to exist
 - B) such validity is really part of construct validity
 - C) it is a problem when external criteria are used
 - D) can be easily eliminated through factor analysis
15. In developing a test
 - A) we need not be concerned with bias at that stage
 - *B) there are several steps that can be taken to eliminate bias
 - C) there is very little to be done to control bias
 - D) bias is not usually a problem at this stage
16. A test item that has different difficulty rates in different racial groups
 - A) is definitely a biased item
 - B) is biased if the item-criterion correlation is low
 - *C) is not necessarily a biased item
 - D) is precisely the type of item needed for cognitive tests

17. The APA Committee's definition of bias emphasizes
 - A) slope rather than intercept bias
 - B) examiner and language bias
 - *C) predictive validity
 - D) test-retest reliability
18. In terms of a regression line, bias is present when:
 - A) neither slope bias nor intercept bias are present
 - *B) predicted criterion scores are consistently too high or low for members of a subgroup
 - C) a test is not valid for the majority group, but appears to be valid for the minority sample
 - D) bias is due to the client or the situation
19. When we take a broad view of test bias, test bias can be defined as:
 - A) intercept bias
 - *B) lack of validity
 - C) differential validity
 - D) unfair advantage for one group
20. From a narrow perspective, bias is evident when:
 - A) the notion that "all people are created equal" is rejected
 - B) a test is unduly influenced by social desirability
 - *C) a minority group has a lower test mean
 - D) slope bias occurs
21. An example of slope bias:
 - A) on test X the mean for whites is 50 and for Hispanics 42
 - B) for a sample of Hispanics the reliability on Test X is .85
 - *C) the SAT correlates .58 with GPA in whites but only .13 for blacks
 - D) scores on the K-ABC do not correlate with scores on Test X
22. When slope bias occurs, the literature suggests that
 - *A) it is the result of poor experimental procedure
 - B) the particular test should be restandardized
 - C) both minority and majority groups will do well on the criterion
 - D) the chosen criterion is simply not appropriate
23. An example of under prediction on a regression equation:
 - A) whites score 500 while Orientals score 620
 - B) the number of false negatives does not equal false positives
 - C) the predicted GPA correlates .86 with the actual GPA
 - *D) the actual GPA is 2.98 while the predicted gpa is 1.59
24. Intercept bias means that
 - *A) the majority group has a higher mean, but both groups do equally well on the criterion
 - B) the minority group has a higher mean
 - C) the majority group does better on the criterion
 - D) the test and the criterion do not correlate above .30
25. In intercept bias, the predictive validity coefficients
 - A) are significantly different in one sample than the other
 - B) are either very low or very high
 - *C) are approximately equal in each sample
 - D) are of the same magnitude as construct validity
26. Studies of intercept bias have shown that
 - A) intercept bias can be nullified by using regression equations
 - *B) if there is a bias, it may be in favor of the minority group
 - C) such bias is present with Hispanics but not blacks
 - D) such bias can be eliminated by using difficult items

27. In regards to selection strategy, the _____ model defines fairness as identical regression lines.
- A) constant ratio
 - B) conditional probability
 - C) equal probability
 - *D) regression
28. If we were to _____, this would be an example of using the regression model as a selection strategy.
- A) admit the same proportion of minority as majority
 - *B) admit highest scoring on SAT regardless of race
 - C) use a test only if it has a high degree of sensitivity
 - D) focus on predictive value as our criterion
29. If we looked at not only the regression equation, but also at the proportion of applicants to be admitted from majority and minority groups, we would be using the _____ model as a selection strategy.
- *A) constant ratio
 - B) conditional probability
 - C) equal probability
 - D) regression
30. In regards to “nature or nurture,” the text suggests that
- A) this is a central question in eliminating bias
 - *B) this is not a good question to ask
 - C) the evidence clearly indicates that intelligence is hereditary
 - D) environmental and family aspects have been neglected
31. The Coleman report showed that differences in mean test scores between minority and majority members were due to:
- *A) socioeconomic status
 - B) differences in school curricula
 - C) the bias present in cognitive tests
 - D) hereditary aspects of intelligence
32. In regards to the argument that more minority children are placed in special education classes, the evidence shows that
- *A) black and low socioeconomic class children are less likely to be placed in such classes
 - B) indeed more minority children are placed in such classes, primarily due to their low intelligence
 - C) such placements are due to tests that show intercept bias
 - D) such placements occur in areas of the US where there is substantial racial prejudice
33. In the study by Poteat et al. (1988) of black and white children referred for special education evaluations and tested with the WISC-R:
- *A) there was no evidence of differential validity of the WISC-R
 - B) white children did substantially better on the WISC-R but not on GPA
 - C) there were significant differences in the regression line
 - D) WISC-R scores did not correlate with GPA
34. The approach called “differential item functioning” focuses on:
- A) how the items relate to the construct validity of the test
 - *B) computing item difficulty separately for minority and majority groups
 - C) how slope bias and intercept bias are confounded
 - D) having expert minority group members review test items

35. Using minority group members as experts to judge which test items are biased:
- A) seems a very useful procedure since their judgments are valid
 - B) yields valid judgments but the interrater reliability is low
 - C) appears useful in personality tests but not cognitive measures
 - *D) such judgments show little relationship to actual empirical findings
36. Cultural bias in test items is usually related to
- *A) the specificity of item content
 - B) intercept bias rather than slope bias
 - C) the use of the constant ratio model
 - D) the degree of acculturation
37. The textbook argues that racial norming
- A) is based on the assumption that minority members score lower on the test
 - *B) assumes that the minority group will do less well on the criterion
 - C) should be used in most situations to offset bias
 - D) is a limited solution to a complex problem
38. As a general conclusion:
- *A) studies of intelligence tests do not show minority bias
 - B) even well designed cognitive tests are somewhat biased
 - C) Mexican-American children do better on verbal than performance items
 - D) tests like the WISC-R predict less accurately minority children's scores on standardized achievement tests
39. Culture-fair tests tend to use _____ items.
- A) abstract reasoning
 - B) unfamiliar
 - *C) non-verbal
 - D) pantomime
40. Establishing the validity of culture-fair tests is problematic because:
- A) bilingual children are not that available
 - *B) most criteria are not culture-free
 - C) most culture-fair tests have poor reliability
 - D) psychological constructs are not equivalent in different cultures
41. In the back translation method we would at a minimum require:
- A) one bilingual examiner
 - B) two people from different cultures but each monolingual
 - C) a bilingual client and a bilingual examiner
 - *D) two bilingual persons
42. Decentering differs from the back translation method:
- A) as intercept bias differs from slope bias
 - *B) as a two way street differs from a one way street
 - C) in that it requires only one bilingual translator
 - D) in that it is a complex procedure to show cultural invariance
43. Etic studies
- A) compare different variables within one culture
 - B) compare different variables but disregard cultural differences
 - *C) compare the same variable across cultures
 - D) use the back translation method but not decentering
44. The study by Laosa (1993) of school readiness in Chicano children
- *A) was given as an example of an etic study
 - B) found that school readiness was not related to parental characteristics
 - C) gave good evidence that tests like the WISC-R are racially based
 - D) illustrates the flaws of the emic approach

45. A finding of the Hu and Oakland (1991) survey of test experts:
- *A) the least developed nations used the fewest tests while socialist nations used the most
 - B) personality measures were the most common type of test
 - C) reliability and validity studies were available for most measures
 - D) the majority of tests were of American origin
46. A finding of the Hu and Oakland (1991) survey of test experts:
- A) achievement tests were used only in industrialized countries
 - B) most achievement tests were of American origin
 - C) most achievement tests had low reliability
 - *D) achievement tests tended to be developed locally
47. _____ refers to changes in behavior and values that occur in minority individuals as they are exposed to the mainstream cultural patterns.
- *A) Acculturation
 - B) Back translation
 - C) Decentering
 - D) Culture-fair
48. The major dimension underlying acculturation:
- A) cultural achievement
 - B) preference for ethnic food
 - *C) language use
 - D) time spent in majority culture
49. A third generation Mexican-American is one born in the U.S. with:
- *A) parents born in the U.S.
 - B) both parents and grandparents U.S. born
 - C) both parents and grandparents born outside the U.S.
 - D) parents born outside the U.S.
50. Acculturation scales typically use one or more of three types of items:
- A) values, attitudes, and language
 - B) language use, proficiency, and preference
 - *C) linguistic, psychological, and sociocultural
 - D) true-false, multiple choice, and Likert
51. An example of a sociocultural item as used in an acculturation scale might be:
- *A) how many brothers and sisters do you have?
 - B) do you prefer to speak Spanish at home?
 - C) what is celebrated on May 5th?
 - D) do you think it is important to honor the elderly?
52. Originally, the items of the Marin acculturation scale
- A) were selected from a larger pool of items through statistical analysis
 - *B) came from other acculturation scales
 - C) were the result of a factor analysis
 - D) were written by the author with consultation of experts
53. The Marin acculturation scale was factor analyzed and the results
- *A) showed a similar factor structure in Hispanics and Anglos
 - B) indicated one major factor of "language preference"
 - C) indicated one major factor of generational distance
 - D) showed a different factor structure for Hispanics than for Anglos

54. If we wanted to validate an acculturation scale, we might validate it against the criterion of:
- A) intelligence test scores
 - *B) generational distance
 - C) educational attainment
 - D) scores on a test like the D-48
55. The Olmedo et al. (1978) Acculturation scale differs from the Marin scale:
- *A) the Olmedo scale was originally longer
 - B) the Marin scale basically measures language
 - C) the Olmedo scale is designed for children
 - D) the Marin scale has been factor analyzed
56. The ARSMA differs from most other acculturation scales in that:
- A) its major factor is not language preference
 - B) it has not been cross-validated
 - C) it is quite long and comprehensive
 - *D) it attempts to measure biculturalism
57. The SL-ASIA developed for Asians
- *A) is modeled on the ARSMA
 - B) is not really an acculturation scale
 - C) seems to have low internal consistency
 - D) is a good example of the back translation method
58. The major factor on the SL-ASIA:
- *A) language preference
 - B) generational identity
 - C) ethnic pride
 - D) food preference
59. Scores on the SL-ASIA correlate significantly with such variables as:
- A) preference for speaking Spanish at home
 - B) GPA in college classes
 - *C) total years living in the U.S.
 - D) religious preference for Buddhism
60. The Cattell Culture-Fair Intelligence Test was
- A) a measure of crystallized intelligence only
 - B) developed through the back-translation method
 - C) heavily verbal in content
 - *D) presumed to be a measure of "g"
61. In developing a culture fair test we would want to emphasize:
- *A) fluid intelligence
 - B) crystallized intelligence
 - C) both fluid and crystallized
 - D) neither fluid nor crystallized
62. In the Smith, Hays, and Solway (1977) study of the Cattell Culture Fair Test with juvenile delinquents:
- *A) whites scored higher on both the Cattell and the WISC-R
 - B) minority delinquents scored higher on the Cattell
 - C) the Cattell and the WISC-R did not correlate with each other
 - D) the Cattell was a better predictor of school achievement

63. The Raven's Progressive Matrices is based on:
- A) Cattell's theory of fluid and crystallized intelligence
 - B) Guilford's structure of intellect model
 - C) Vernon's hierarchical model
 - *D) Spearman's two factor theory
64. The Raven's Progressive Matrices yields a(n):
- *A) total score
 - B) Verbal and Performance IQ
 - C) index of estimated learning potential
 - D) ratio IQ
65. In regards to gender differences on the Raven's:
- A) gender differences on the Standard form only
 - B) gender differences in adolescents but not adults
 - *C) no gender differences on any of the three
 - D) gender differences on all three forms
66. The Raven seems particularly useful with groups where
- A) visual acuity is impaired
 - B) fatigue is a major testing problem
 - C) social desirability may be a concern
 - *D) language processing needs to be at a minimum
67. In England, the Standard Progressive Matrices is often administered with the:
- *A) Mill Hill Vocabulary Scale
 - B) Differential Aptitude Test
 - C) K-ABC
 - D) SOMPA
68. The Standard Progressive Matrices can be characterized as:
- A) an individual test with a 20 minute time limit
 - B) requiring a skilled administrator
 - *C) untimed and group administered
 - D) a group test for children ages 5 to 12
69. In regards to norms on the Standard Progressive Matrices:
- A) extensive but only for children
 - *B) extensive, for a variety of groups in different cultures
 - C) rather limited, primarily British samples
 - D) based on samples of convenience, primarily Canadian
70. We would expect scores on the Standard Progressive Matrices to correlate with scores on the Stanford-Binet or WISC in the range:
- A) .00 to .10
 - B) .10 to .30
 - C) .30 to .50
 - *D) .50 to .80
71. The Standard Progressive Matrices is said to have a "restricted ceiling." This means that:
- A) very high scores are impossible to obtain
 - B) the decision to discontinue testing after 3 failures is too harsh
 - *C) the test may be too easy for a number of subjects
 - D) the test is composed of too many difficult items
72. The Coloured Progressive Matrices is:
- A) specifically designed for minority groups
 - B) an easy form for use with retarded subjects
 - *C) similar to the Standard but with colored items
 - D) intended for use with gifted children

73. The Advanced Progressive Matrices
- *A) consists of two sets of items, one of which can be administered with or without time limit
 - B) is a more difficult version of the Standard PM but with no time limit
 - C) was developed as an alternate form to the Coloured PM
 - D) has excellent norms especially for U.S. children
74. The items on the D-48 consist of
- A) verbal analogies
 - B) geometric designs
 - *C) sequences of dominoes
 - D) symbol completion items
75. The relative difficulty level of D-48 items is:
- A) a function of the age of the subject and language spoken
 - *B) constant for different age groups in different countries
 - C) about .80, as expected of a difficult test
 - D) good evidence that the test is biased against minorities
76. Most of the validity information for the D-48 is of the _____ type.
- *A) concurrent
 - B) predictive
 - C) construct
 - D) discriminant
77. The D-48 was discussed in the textbook because:
- A) unlike the Raven, the D-48 was developed in the U.S.
 - B) it is an excellent example of Guilford's theory
 - C) like the SOMPA, it reflects a pluralistic perspective
 - *D) it is widely known, except in the U.S.
78. Gough and Domino (1963) list five reasons why professionals should pay attention to the D-48. Which is NOT a reason:
- A) the D-48 is almost entirely nonverbal
 - *B) uses novel stimuli unknown in most cultures
 - C) more highly loaded on "g" than the Raven's
 - D) easy to administer and score
79. The SOMPA is a:
- A) structured interview
 - B) theoretical model
 - *C) battery of tests
 - D) statistical analysis
80. The SOMPA reflects a tripartite model of three perspectives:
- A) psychological, social, and physical
 - B) verbal, performance, and pluralistic
 - C) minority, majority, and integrative
 - *D) medical, social system, and pluralistic
81. As part of the medical perspective, the SOMPA utilizes the
- *A) Bender Visual Motor Gestalt Test
 - B) Luria-Nebraska Children's Neuropsychological Test Battery
 - C) Adaptive Behavior Inventory for Children
 - D) WISC-R

Testing in a Cross-Cultural Context

82. The purpose of the Adaptive Behavior Inventory for Children is to:
- A) observe how a child behaves in the classroom
 - *B) evaluate a child's competence, from parental information
 - C) assess a child's classroom skills, from teacher observations
 - D) obtain a brief medical/health history of the child
83. In the SOMPA, the aim of the pluralistic perspective is basically to:
- *A) determine how a child's world differs from the Anglo culture
 - B) estimate how the child meets specific cultural demands
 - C) integrate the various sources of information into one
 - D) identify the proportion of variance attributable to culture
84. In using the WISC-R, the SOMPA
- A) uses national norms
 - B) disregards ethnic background
 - C) uses only the Performance subtests
 - *D) uses racial norming
85. As far as the SOMPA is concerned, we can conclude that:
- *A) there are simply too many problems to make it work
 - B) this is an exciting new approach with solid validity
 - C) the criticisms are few and can be easily corrected
 - D) in a few years it should be as popular as the Stanford-Binet
86. The Black Intelligence Test of Cultural Homogeneity was standardized on:
- A) both white and black police department applicants
 - B) a random sample based on census characteristics
 - *C) a sample of black high school students
 - D) black children and adolescents
87. We can conclude that the BITCH-100 is
- A) a valid measure of intelligence especially with blacks
 - B) a valid measure of cultural perspective
 - C) a valid predictor of college GPA for blacks
 - *D) is not a valid measure of intelligence even for blacks
88. The Johnson and McGowan (1984) study of the school performance of low income Mexican-American children showed:
- A) that the role of the father as a model is highly important
 - *B) a modest relationship between intellectual functioning and school achievement
 - C) significant gender differences in school achievement
 - D) that the Stanford-Binet and WISC-R are worthless as grade predictors in minority children
89. On The SAT, in regards to Hispanic students, we can conclude that:
- *A) the SAT has equal predictive validity for Hispanics and for Anglos
 - B) the SAT is clearly biased and should not be used
 - C) in general, the predictive validity of the SAT is excellent
 - D) their mean scores are not statistically different from those of whites
90. In the regression equation developed by McCornack (1983) to predict first-semester college GPA, the most important variable was:
- A) scores on an SAT composite
 - *B) high school GPA
 - C) parental support
 - D) lack of financial worries

91. ETS reviewed almost 500 studies of the GRE. Typical correlations between GRE scores and graduate grades ranged between:
- A) .00 and .18
 - *B) .15 and .31
 - C) .46 and .53
 - D) .62 and .81
92. In the Whitworth and Barrientos (1990) study of Hispanic and Anglo graduate students, the results indicated that:
- *A) GRE scores did not predict graduate grades for either group
 - B) the GRE was the best predictor of grades for Anglo students only
 - C) restriction of range had no influence
 - D) degree of motivation was the crucial variable
93. We can conclude that on the MMPI
- *A) if there are racial differences, they are related to personality and/or behavior
 - B) there is substantial evidence that the MMPI is biased
 - C) both slope bias and intercept bias are present in the MMPI
 - D) blacks typically score lower than whites, but only 1 or 2 points
94. Using the regression equation developed by McCornack (1983) to predict first semester GPA, when predicted GPA was compared to actual GPA the results indicated
- A) almost no correspondence between the two
 - *B) a fair degree of correspondence, with r 's in the .30 to .50 range
 - C) that first semester GPA correlates significantly with later GPA
 - D) SAT scores were a better predictor than high school GPA

12 Disability and Rehabilitation

1. In the national surveys, it was found that some _____ million individuals in the U.S. were considered disabled.
A) 2
B) 15
*C) 27
D) 46
2. In the national surveys, of those found to be disabled approximately _____ were not working.
*A) 2 out of 3
B) 1 out of 10
C) 9 out of 10
D) 1 out of 3
3. The concept of “work sample” involves:
A) a person’s resume of previous experience
*B) a simulated work activity
C) a statistical model based on sampling distribution
D) observing an individual at work
4. In tests related to diagnostic status, longitudinal studies of disabled children find:
A) rather poor stability, as in studies of normal children
B) excellent test-retest reliability but poor internal consistency
C) the same results as with normal children
*D) substantial stability, unlike studies of normal children
5. In testing a rehabilitation client it is important to
*A) know as much as possible about the client’s abilities and limitations
B) know as little as possible about the client so evaluation is objective
C) realize that most blind clients can read braille
D) utilize speeded power tests, as these are the most valid
6. In testing a blind client the examiner should
A) pet and speak to the seeing-eye dog in order to establish rapport
B) avoid phrases like “it’s nice to see you”
*C) let the blind client take the examiner’s arm
D) speak only through an interpreter
7. In testing handicapped individuals, the literature suggests a “multivariate” approach. This means:
A) retesting the client several times to be sure scores are correct
B) having different examiners administer different tests
C) doing an analysis of variance on the results
*D) using a battery of instruments

8. In doing diagnostic assessment the first step is to:
 - A) interview the client to determine his/her needs
 - *B) determine what the goals of assessment are
 - C) obtain the background medical information
 - D) select the appropriate instruments
9. Amanuensis:
 - *A) a person who marks the answers for the test taker
 - B) a trained examiner who can use sign language
 - C) a client who is visually disabled
 - D) a special test procedure used with deaf persons
10. In testing a blind client, a chart can be presented as an embossed figure:
 - A) so the item still measures the same functions as in a sighted client
 - *B) but the item may now be measuring different functions than the original
 - C) the literature indicates such items are equivalent to the original
 - D) the literature indicates that only 20% of the blind can “read” such charts
11. In regards to time limits, the ideal solution in testing a disabled client is:
 - A) to use only power tests
 - B) to give clients as much time as they need
 - *C) not to use tests with time limits
 - D) to use the time limit since disabled clients do not need extra time
12. In testing clients with disabilities, tests can be modified in three major ways. Which one does NOT belong?
 - A) using braille or audiotapes
 - B) using tests that do not have a time limit
 - C) omitting specific test questions
 - *D) using an examiner of the same race as the client
13. The Panel on Testing of Handicapped People recommend that:
 - A) on standard measures like the WISC there was no need for further validity data
 - *B) validity information on disabled samples was needed
 - C) tests of intelligence not be used with disabled clients
 - D) where possible, examiners with some disability be used
14. The major conclusion of the ETS study of performance of individuals with disabilities on the SAT and the GRE:
 - A) there is no need for special test accommodations for these clients
 - *B) non-standard versions of these tests were comparable to the original
 - C) the non-standard versions of the SAT and GRE are less valid
 - D) these tests are simply not appropriate for such clients
15. Both the Strong and the Kuder are of limited use with some disabled clients because:
 - A) there are no norms on disabled clients
 - B) most disabled clients are not seeking such information
 - C) the test results are of little use to clients who will not work
 - *D) they both involve the ability to read at a fairly high level
16. The items on the Geist Picture Interest Inventory Revised are drawings of:
 - *A) work activities
 - B) tools used in different occupations
 - C) recreational activities
 - D) conflict situations

17. The Geist Picture Interest Inventory - Revised is designed for:
 - A) disabled individuals who have visual or hearing disabilities
 - *B) normal adolescents of high school age and adults with limited verbal skills
 - C) normal or disabled children younger than 16
 - D) adults with limited visual acuity
18. Among the criticisms of the Geist Picture Interest Inventory - Revised:
 - A) the factor structure yields one main factor
 - B) the theory behind the test is now outdated
 - C) the test is too complicated for many disability clients
 - *D) many occupations reflect gender stereotypes
19. In general, most picture interest inventories
 - A) are normed on normal children
 - *B) are technically inadequate
 - C) are very useful with rehabilitation clients
 - D) have excellent reliability but poor validity
20. Career interest test results of disabled clients may be invalid because:
 - A) such tests are typically normed on normal samples
 - *B) such clients do not have the type of experiences needed for stable career interests
 - C) Holland's theory does not take into consideration developmental aspects
 - D) they are essentially speeded power tests
21. In regards to college admissions testing of disabled students, the limited evidence suggests that:
 - A) tests like the ACT and the SAT should not be used with disabled students
 - B) scores on tests like the SAT are excellent predictors of subsequent performance
 - *C) there is no difference in prediction of grades for disabled and normal students
 - D) disabled students should not be required to take such tests
22. Most experts define _____ as the result of life long learning.
 - A) achievement
 - B) intelligence
 - C) interests
 - *D) aptitudes
23. A major conclusion of Ghiselli's (1966) review of occupational aptitude tests:
 - A) most such tests had little applicability to rehabilitation clients
 - *B) tests possessed sufficient predictive power to be of value in personnel selection
 - C) proficiency as measured by ratings was not a useful criterion
 - D) tests of perceptual accuracy were the most valid
24. In Ghiselli's (1966) review of occupational aptitude tests, criteria were divided into two major categories:
 - *A) training and proficiency
 - B) aptitude and achievement
 - C) intellectual abilities and personality
 - D) perceptual accuracy and motor abilities
25. Minimum competency testing typically refers to:
 - *A) high school graduation
 - B) disabled students applying to college
 - C) minority high school students
 - D) all adults with disabilities

26. As far as the question of which norms to use on a test that we administer to a client with disabilities:
- A) the best norms are national norms
 - *B) there often is no available data
 - C) we should use whatever norms best match the client
 - D) we should not use norms but only raw scores
27. The General Aptitude Test Battery is the most widely used test for
- A) assessing deaf and mentally retarded persons
 - *B) hiring in government agencies
 - C) measuring intelligence of rehabilitation clients
 - D) determining the aptitudes of high school students
28. The General Aptitude Test Battery was originally developed by
- A) assessing military personnel in World War II
 - B) studying different occupations to see what skills were needed
 - C) comparing normal and disabled subjects on various tasks
 - *D) factor analyzing some 50 different tests
29. On the GATB the ability to proofread correctly and to perceive pertinent details is defined as:
- A) finger dexterity
 - B) spatial aptitude
 - C) general learning ability
 - *D) clerical perception
30. The finger dexterity and manual dexterity portions of the GATB differ from all other aptitudes in that:
- *A) they are assessed through performance tests
 - B) they use true-false items exclusively
 - C) these tests were chosen because of factor validity
 - D) their internal consistency is substantially lower
31. The development of the GATB was guided primarily by
- A) occupational theories
 - *B) factor analysis
 - C) construct validity
 - D) hierarchical theories
32. One of the unique aspects of the GATB:
- A) routine use of racial norming
 - *B) occupational aptitude patterns
 - C) little technical material available
 - D) can be administered and scored by the client
33. In regards to testing disabled individuals, the Standards for Educational and Psychological Testing recommends that
- A) norms based on disabled samples be used
 - B) norms for the general population be used
 - *C) either general norms or specific norms be used, depending on the purposes of testing
 - D) norms be disregarded as they can be misleading when testing disabled individuals
34. "Titles" of the Civil Rights Act of 1964 are simply:
- *A) provisions for implementing various aspects
 - B) sections of the Act numbered for easy identification
 - C) penalties associated with various illegal acts
 - D) general principles that apply to testing of minorities

35. Under the 1973 Rehabilitation Act disabled individuals
- A) may not be tested psychologically
 - B) cannot be given intelligence tests
 - C) must be admitted to postsecondary education
 - *D) shall not be subjected to discrimination
36. The various guidelines of Section 504 of the Rehabilitation Act of 1973 can best be summarized by:
- A) a complete evaluation must be carried out before a child is assigned to a special education program
 - B) “flagging” a person’s test scores is illegal
 - C) it is illegal to discriminate against any person
 - *D) tests must not discriminate and must reflect the intended construct rather than the disability
37. The Education for All Handicapped Children Act mandates that a hearing impaired child must be administered tests
- A) in braille only
 - *B) in the “normal” mode of communication
 - C) in any appropriate minority language like Spanish
 - D) using the majority language only
38. One of the major problems in testing the visually impaired:
- A) most clients refuse to be tested
 - *B) most tests involve sight
 - C) norms for the visually impaired can be misleading
 - D) most examiners have had little experience with braille
39. Macular degeneration
- A) involves clouding of the lens of the eye
 - B) would be classified as congenital blindness
 - *C) primarily affects the elderly
 - D) is typically associated with severe retardation
40. In general, _____ seem to be valid predictors of work or academic success for the visually impaired.
- *A) verbal cognitive tests
 - B) tests like the Wechsler series
 - C) interview ratings
 - D) self ratings of adjustment
41. The Minnesota Rate of Manipulation Test was discussed primarily to point out that
- A) some tests are excellent measures of visual acuity
 - B) the cognitive abilities of the visually impaired can be assessed
 - C) tests with speeded time limits are invalid for the visually impaired
 - *D) norms based on sighted individuals may not be appropriate
42. In regards to the cognitive abilities of the visually impaired:
- A) they can be validly assessed by a test like the WAIS-R
 - *B) verbal portions of the Wechsler scales are particularly useful
 - C) the evidence suggests that such clients score below the mean
 - D) there is substantial evidence for the validity of cognitive tests
43. The two forms of the Perkins-Binet are
- A) for those who are congenitally blind versus adventitiously blind
 - *B) for those with no usable vision versus some usable vision
 - C) essentially identical except that one form uses time limits
 - D) of use in the assessment of test-retest reliability

44. In general we can conclude that the Perkins-Binet
 - A) is a highly useful instrument for the visually impaired
 - B) is an alternate form of the Stanford-Binet
 - *C) is psychometrically inadequate and probably lacks validity
 - D) can be used instead of the verbal subtests of the WISC
45. Haptic abilities involve
 - *A) active tactual perception
 - B) long-term memory
 - C) verbal abilities
 - D) alpha reliability
46. The Dutch cognitive test discussed in your textbook was used to illustrate
 - A) how available tests can be modified for use with the visually impaired
 - B) how difficult it is to use tests developed in a different culture
 - C) the predictive utility of haptic abilities for the visually impaired
 - *D) the development of standardized tests for the visually impaired
47. In the Dutch cognitive test, dark or light gray test materials were used
 - A) to maximize any vision a child may have
 - B) so that the child can indicate whether he/she has some vision
 - *C) to minimize any vision a child may have
 - D) to control for any differences in vision with sighted children
48. In the study of the tactual form of the D-48, the results suggested that the test
 - A) had low, unacceptable reliability
 - *B) is quite valid with blind adults
 - C) did not correlate significantly with WAIS Verbal IQs
 - D) was too easy for most visually impaired subjects
49. The _____ position views visual perception as genetically determined and basically a passive process.
 - *A) nativist
 - B) empiricist
 - C) haptic
 - D) psychoanalytic
50. The Frostig test essentially measures
 - A) nonverbal intelligence of visually impaired children
 - B) the degree of school difficulties a blind child experiences
 - *C) the presence of difficulties in processing visual stimuli
 - D) five separate factors of visual perception
51. In regards to the reliability of the Frostig:
 - A) there is no difference between psychologists and others
 - *B) trained examiners obtain higher coefficients
 - C) the reliability of the individual subtests is quite high
 - D) the reliability of the total test is inadequate
52. Factor analytic studies of the Frostig
 - A) support the author's theory of visual perception
 - B) show different factors with sighted and normal children
 - *C) point to one single factor
 - D) typically show five major dimensions
53. An infection of the membranes that surround the brain:
 - *A) meningitis
 - B) rubella
 - C) German measles
 - D) lingual deafness

54. A person who becomes deaf at about age 45 would most likely be labelled as:
- A) prelingually deaf
 - B) postlingually deaf
 - C) prevocationally deaf
 - *D) late deafened
55. In testing hearing impaired persons, it is important to make sure that
- A) they are fluent in American Sign Language
 - B) excellent rapport be established
 - *C) they understand the test instructions
 - D) their hearing impairment is not aggravated by other conditions
56. Some of the challenges of testing the hearing impaired can be met by using _____ items.
- *A) non-verbal
 - B) haptic
 - C) verbal
 - D) spatial
57. In general, when testing the hearing impaired:
- A) the presence of a sign language interpreter is unnecessary
 - *B) group testing is not valid
 - C) the use of braille is a useful procedure
 - D) group testing can allay the client's fears
58. Bragman's (1982) list of 22 tests commonly used with the hearing impaired covers tests primarily developed for:
- *A) hearing individuals
 - B) the hearing impaired
 - C) mentally retarded individuals
 - D) the visually impaired
59. In regards to test norms for the hearing impaired:
- A) most major tests include these
 - *B) few tests commonly used with the hearing impaired have them
 - C) they are not necessary for most tests
 - D) they are generally not useful for individual assessment
60. The literature suggests that the 16PF may
- A) be quite useful if administered by American Sign Language
 - B) yield high social desirability scores with the hearing impaired
 - C) be very useful since it has hearing impaired norms
 - *D) not be valid or reliable with the hearing impaired
61. In assessing the hearing impaired, the most frequent modification of standard tests involves:
- *A) changes in test instructions
 - B) testing the limits
 - C) disregard of any time limits
 - D) the use of same group norms
62. In testing the hearing impaired, alternate forms of tests can be particularly useful because:
- A) alternate form reliability is particularly crucial
 - B) hearing impaired clients often need to be retested
 - *C) they provide items that can be used as examples
 - D) both forms can be administered to increase reliability

63. There is some evidence that hearing impaired children have less well developed _____ than those of hearing children.
- *A) schemas
 - B) motor coordination
 - C) visual cortex areas
 - D) all of the above
64. In Gibbins' (1989) survey of school psychologists who serve the hearing impaired, the most popular measure of cognitive abilities was the:
- A) Self-Concept Scale for the Hearing Impaired
 - *B) WISC-R Performance Scale
 - C) MMPI
 - D) Vineland Adaptive Behavior Scale
65. In Gibbins' (1989) survey of school psychologists who serve the hearing impaired, the most commonly used test to assess adaptive behavior was the
- A) Wide Range Achievement Test
 - B) Hiskey-Nebraska Test of Learning Aptitude
 - C) Samuelson Test of Adaptive Behavior
 - *D) Vineland Adaptive Behavior Scale
66. The popularity of projective drawings in the assessment of the hearing impaired is probably due to:
- A) research findings of excellent validity
 - *B) their ease of administration
 - C) availability of specific norms for the hearing impaired
 - D) development of special forms for this population
67. The Stanford-Binet is not generally recommended for use with the hearing impaired because of its
- A) stringent time limits
 - B) lack of items that can be used as examples
 - C) norms that are based on hearing children
 - *D) heavy emphasis on language ability
68. In general, the research results suggest that the nonverbal intelligence of the hearing impaired
- *A) does not differ qualitatively from that of normal hearing
 - B) is typically 10 to 15 points below that of normal hearing
 - C) is much more concrete and less creative
 - D) differs substantially in its factor structure
69. A "typical" group profile for hearing impaired on the Wechsler:
- *A) low on Coding, well on Picture Completion, and above average on Block Design
 - B) low on Block Design, Coding, and Picture Completion
 - C) above average on Object Assembly and on Coding, and low on Picture Completion
 - D) high on Picture Arrangement and on Coding
70. In the Hirshoren et al. (1977) study of prelingually deaf children:
- A) the pattern of intercorrelations for WISC-R subtests indicated three basic factors
 - B) the mean IQ on the WISC-R Performance Scale was 100
 - *C) scores on the WISC-R Performance scale correlated substantially with the Hiskey-Nebraska
 - D) neither the WISC nor the Hiskey-Nebraska were shown to be valid

71. The _____ is one of the few standardized tests for which there is a standardized version for deaf children.
- A) Stanford-Binet
 - B) Raven Progressive Matrices
 - C) Hiskey-Nebraska
 - *D) WISC-R Performance Scale
72. In the Forde (1977) study of the Peabody Picture Vocabulary Test with the hearing impaired, the results showed:
- A) low internal consistency and inadequate validity
 - *B) substantial correlations with the Stanford Achievement Tests
 - C) a factor structure consonant with the two factor theory
 - D) marginal correlations with the WISC-R Performance Scale
73. Studies with the special edition of the Stanford Achievement Test show that hearing impaired students aged 20 or above have
- A) excellent math computation skills
 - B) skills equal to those of hearing students
 - C) visual problems that interfere with their performance
 - *D) poor reading skills
74. The test-retest reliability of the special edition of the Stanford Achievement Test:
- *A) seems to be excellent even over a 5 year period
 - B) is adequate for brief time periods only
 - C) is marginal at best
 - D) has not been investigated with hearing impaired
75. The Farrugia (1982) study of the vocational interests of the hearing impaired found that hearing impaired
- A) had unrealistic vocational aspirations
 - *B) preferred manual activities over verbal learning areas
 - C) showed high interests in art and drama
 - D) had identical interest patterns to those of hearing persons
76. In using self-concept measures with the hearing impaired, there are at least two concerns:
- A) marginal reliability and strict time limits
 - *B) reading level and inappropriate items
 - C) changes in instructions and lack of norms
 - D) lack of appropriate norms and poor validity
77. In developing the Self-Concept Scale for the Hearing Impaired, items were chosen from an item pool on the basis of:
- A) high test-retest reliability
 - B) factor analysis
 - C) significant correlations with a criterion
 - *D) none of these
78. One of the major limitations of the Self-Concept Scale for the Hearing Impaired:
- A) the 30 minute time limit is too stringent
 - *B) poor correlations with teacher and/or professional ratings
 - C) poor test-retest reliability for the total score
 - D) inappropriate response choices
79. In the study by Ouellette (1988) who administered the House-Tree-Person to the hearing impaired, the inter-rater reliability was:
- *A) quite low with a median of about .32
 - B) substantial for four of the seven scales
 - C) adequate for psychologists but not for counselors
 - D) low, but due to a statistical artifact

80. The use of behavioral rating scales with the hearing impaired:
- A) is generally limited due to reading limitations
 - *B) can be quite useful
 - C) is generally not useful because of low inter-rater reliability
 - D) has not been investigated
81. The Bolton and Brown (1971) study using the Semantic Differential as a measure of attitudes toward work for use with the hearing impaired, illustrated:
- A) the flexibility of the Semantic Differential
 - B) how high reliability can be achieved by careful work
 - *C) the lack of validity data in this area
 - D) the use of factor analysis in test development
82. A stoma is essentially a
- *A) surgically constructed passageway for body waste
 - B) multipluralistic testing approach applied to physically disabled
 - C) a type of T score with a mean of 60
 - D) a personality test for spinal cord injury patients
83. Convergent validity of the Ostomy Adjustment Scale was reflected by:
- A) insignificant correlations with social desirability
 - *B) positive correlations with various aspects of surgery
 - C) negative correlations with measures of self-esteem
 - D) positive correlations with self ratings of adjustment

13 Testing in the Schools

1. The general objective of assessment in preschool educational settings is to:
 - *A) make appropriate decisions to facilitate psychoeducational development
 - B) identify children with developmental disabilities
 - C) determine the eligibility of children for special programs
 - D) diagnose children who are either gifted or retarded
2. The “test-teach-test” approach urges the use of:
 - A) norm-referenced tests with good reliability
 - B) assessment procedures that have high test-retest reliability
 - C) interviews rather than actual tests
 - *D) curriculum-based assessment measures
3. One of the most valuable assessment methods to study young children:
 - A) interviews of parents and of teachers
 - *B) direct behavioral observation
 - C) rating scales
 - D) protective techniques
4. In rating scales, error variance can be defined as:
 - A) test-retest reliability
 - B) lack of discriminant validity
 - C) a subcategory of the multi-trait multi-method matrix
 - *D) unwanted variation in scores
5. Projective techniques are of limited use with preschool children because these children lack:
 - A) spatial skills
 - B) fantasy and imagination
 - *C) verbal skills
 - D) well defined emotions
6. Piagetian based scales
 - A) have been developed and are used widely
 - *B) have been criticized and not found widespread use
 - C) are particularly useful as process oriented measures
 - D) are only applicable to personality development
7. Cognitive Test X and Test Y correlate about $+ .92$. We could therefore conclude that a child would most likely obtain identical
 - A) IQ's on the two tests
 - *B) rankings vis-a-vis their peers
 - C) test-retest correlation coefficients
 - D) z scores

8. The main assumption of process oriented assessment approaches is that:
 - A) cognitive skills underlie personality functioning
 - B) the test-teach-test sequence is of limited use
 - *C) what is important is how one arrives at an answer
 - D) the examiner needs to process or understand the child
9. Reliability of tests administered to young children:
 - A) can be low because of limitations of the tests
 - B) is usually quite high since preschool children show individual differences
 - C) is ordinarily satisfactory since measures like the WPPSI are stable
 - *D) can be low because of rapid developmental changes
10. Before testing a preschool child the examiner should
 - A) meet with the child and the parents to establish rapport
 - B) make sure the testing room has attractive furniture
 - *C) have a clear and specific referral question
 - D) review a handbook of child psychology
11. The California Achievement Tests are designed for use in:
 - *A) kindergarten through the 12th grade
 - B) first through fifth grade
 - C) eighth through 12th grade
 - D) kindergarten through sixth grade
12. The California Achievement Tests may be described as:
 - A) individual tests requiring a psychologist as examiner
 - B) essay type tests, group administered
 - C) similar to the WISC in structure and administration
 - *D) a group administered battery with teachers as examiners
13. The three formats of the California Achievement Tests differ:
 - A) in the grade level in which they are to be used
 - *B) in number of items per subtest, and kind of scores reported
 - C) only in superficial ways, so they are basically equivalent forms
 - D) in the specific subtests that are included under each
14. One of the major psychometric problems of the California Achievement Tests is that:
 - A) internal consistency coefficients tend to be low
 - B) the reliability is low at younger age levels
 - *C) the subtests intercorrelate substantially
 - D) little evidence is available on its validity
15. In achievement test batteries the emphasis is usually on _____ validity.
 - A) construct
 - *B) content
 - C) concurrent
 - D) face
16. In regards to the norms of the California Achievement Tests:
 - A) large, but basically samples of convenience
 - B) chosen on the basis of U.S. Census data
 - C) based primarily on California school children
 - *D) large and representative
17. The Lake Wobegon effect refers to:
 - A) more children scoring below average because of outdated norms
 - B) what statisticians would call regression towards the mean
 - C) the altering of children's test scores by teachers
 - *D) a large percentage than expected obtaining above average scores

18. In the Neeper and Lahey (1984) study of the common dimensions of teacher rating scales, the major factor was:
 - A) inattentiveness
 - *B) conduct disorder
 - C) school anxiety
 - D) social skills
19. In high school one of the critical aspects of adolescence is:
 - A) forming an identity
 - B) achieving in academic subjects
 - *C) the development of social competence
 - D) satisfactory sexual experiences
20. The items for the Cavell and Kelley (1994) measure of social competence came from:
 - *A) self reports of adolescents
 - B) the writings of experts
 - C) other available scales
 - D) the authors' own expertise
21. The internal consistency coefficients for the Cavell and Kelley (1994) measure of social competence are somewhat inflated because:
 - A) they were computed on the initial pool of items rather than the final scales
 - B) the subscales are intercorrelated with each other
 - C) both alternate forms were used
 - *D) they were computed on the same sample used for factor analysis
22. In the Cavell and Kelley (1994) measure of social competence, the most frequently occurring situations involved:
 - *A) parents and siblings
 - B) work situations
 - C) conflict with teachers
 - D) problem behaviors
23. The GED tests are used to:
 - A) determine whether a high school student should graduate
 - B) decide on college admission where SAT scores are not available
 - *C) award high school level equivalency
 - D) assess an individual's overall level of social competence
24. On the GED, alternate form reliability is somewhat lower than internal consistency because:
 - A) those who take the GED are more homogeneous than h. s. students
 - *B) different forms were administered on different days
 - C) internal consistency is a more stringent calculation
 - D) both alternate forms were administered at one sitting
25. In general, the reliability of the GED is:
 - A) marginal at best
 - B) in the poor range
 - *C) quite substantial
 - D) too low to make individual decisions
26. The major type of validity applicable to the GED is:
 - *A) content
 - B) construct
 - C) criterion
 - D) predictive

27. The GED is also administered to national samples of high school graduating seniors in order to determine:
- A) whether the GED can be used as a college admissions test
 - B) if high school students are showing minimum competency
 - C) whether available norms are still accurate
 - *D) that test scores reflect a high school equivalency performance
28. Somewhere near _____ of graduating high school seniors would in fact fail the GED tests.
- A) one half
 - *B) one third
 - C) one tenth
 - D) one out of one hundred
29. As illustrated by the GED, we ordinarily would expect _____ validity to be _____ than _____ validity.
- A) concurrent, higher, parallel forms
 - B) predictive, lower, concurrent
 - C) construct, higher, concurrent
 - *D) concurrent, lower, parallel forms
30. Follow up studies of GED graduates indicates that passing the GED leads to various benefits. Such data:
- A) can be considered substantial construct validity
 - B) is limited by the poor reliability of the GED
 - C) can be considered as part of content validity
 - *D) is somewhat suspect since nongraduates were not surveyed
31. The NAEP (National Assessment of Educational Progress) is essentially a(n) _____ test.
- A) personality
 - B) cognitive
 - *C) achievement
 - D) aptitude
32. Not every student is tested with the NAEP (National Assessment of Educational Progress) because:
- *A) the test is designed to measure group characteristics
 - B) only those who are achieving above the median are tested
 - C) it is too long to be administered to all students
 - D) only students who are at risk for failure are tested
33. The results of the NAEP (National Assessment of Educational Progress) are presented as:
- A) total T scores for each individual
 - B) group means for each school
 - *C) percentage of students who answer correctly a particular item
 - D) regression equations to predict academic achievement
34. The review by Bridgeman and Lewis (1994) of multiple choice and essay sections in Advanced Placement exams found that:
- *A) reliability for essay sections was substantially lower
 - B) performance on essay sections was a better predictor of GPA
 - C) multiple choice and essay sections correlated significantly
 - D) multiple choice items were only marginally reliable
35. The findings of the Bridgeman and Lewis (1994) review of Advanced Placement exams are a good argument for:
- A) not using multiple choice items
 - B) doing away with such placement exams
 - C) the idea that gpa is not a reliable criterion
 - *D) eliminating essay questions

36. When the SAT was originally introduced it was intended to:
- A) allow minority students to attend prestigious universities
 - *B) fight discrimination in college admissions
 - C) assess minimum competencies of applicants
 - D) give women an equal opportunity to enter college
37. The SAT quantitative appears to:
- A) have two major dimensions
 - *B) be unidimensional
 - C) have a high loading on g
 - D) be made up of four separate factors
38. In a multiple choice test we can specify what content is to be covered and therefore content validity is:
- A) assumed to be identical with construct validity
 - B) not even considered
 - *C) higher than that of essay exams
 - D) typically higher than its reliability
39. How is test sophistication removed as an extraneous variable? By:
- *A) having easy to read and simple directions and items
 - B) having students take a course on the particular test
 - C) scoring the test with a formula to control for guessing
 - D) doing an analysis of variance to remove such error variance
40. Reading Taking the SAT
- *A) has minimal effect on subsequent test scores
 - B) improves SAT scores by about $\frac{1}{2}$ standard deviation
 - C) is very helpful to students with little test sophistication
 - D) improves SAT scores for minority clients
41. The text says that the SAT gender difference is “somewhat peculiar.” Why? Because
- *A) women do better on the predicted criterion
 - B) men do better on the predicted criterion
 - C) the difference does not fit with the construct validity of the SAT
 - D) male students outperform females in math ability
42. The gender gap on the SAT may be due to:
- A) the fact that fewer women select majors like math or engineering
 - B) different types of courses taken in high school
 - C) the SAT not being a valid predictor of academic achievement
 - *D) the increasing number of women who take the test
43. As a predictor of college success the SAT _____ the performance of _____ .
- A) underpredicts, men
 - B) overpredicts, women
 - *C) underpredicts, women
 - D) overpredicts, both men and women
44. Sheehan and Gray (1991) compared the SAT and a standardized algebra exam and found:
- A) a gender gap on both tests
 - *B) no gender gap on the algebra test
 - C) a gender difference on both the algebra test and GPA
 - D) no gender differences on any of the measures
45. Byrnes and Takahira (1993) suggested that the gender differences on the SAT may reflect:
- A) genetic differences between the sexes
 - *B) male students performing certain cognitive operations more effectively
 - C) women having more difficulties with multiple choice tests
 - D) that women choose less difficult majors

46. If minority members tend to score lower on the predictor than on the criterion as compared to the majority, use of the SAT will lead to:
- A) greater incidence of false positives
 - B) overprediction of gpa for all students
 - *C) greater incidence of false negatives
 - D) a higher rate of errors for both minority and majority
47. A number of studies have shown that in minority students _____ is a better predictor of college achievement than test scores.
- *A) high school achievement
 - B) self report
 - C) participation in sports
 - D) motivation
48. In McCornack's (1983) study of SAT scores of minorities, the regression equation predicted GPA:
- *A) relatively well for all groups
 - B) relatively poorly for all groups
 - C) well for majority but poorly for minorities
 - D) well for both white and Asian students but poorly for others
49. In McCornack's (1983) study of the SAT and minority students, the results showed that high school GPA was a _____ predictor of college GPA _____.
- *A) better; for all students except Blacks
 - B) better; for all students
 - C) poorer; than the SAT
 - D) poorer; than the SAT in Asian students only
50. Goldman and Richards (1974) analyzed SAT scores for both Mexican- and Anglo-American students and concluded that:
- A) regression equations should be developed on the Anglo group only
 - B) regression equations developed on the Mexican-American sample underpredict
 - C) the predictive validity of the SAT as to graduation is very poor
 - *D) the predictive validity of the SAT is similar in both groups
51. James Crouse has recommended that colleges
- A) abandon all admissions testing as invalid
 - *B) use standardized achievement tests rather than the SAT
 - C) continue using the SAT in addition to standardized achievement tests
 - D) use regression equations to predict academic success
52. Crouse's argument about the SAT is based on the notion that the SAT
- A) is simply not a valid procedure
 - B) really assesses hereditary abilities rather than learning
 - *C) provides redundant information
 - D) has substantially lower predictive validity than achievement tests
53. In Crouse's view, the advantage of achievement tests over the SAT is that achievement tests
- *A) would promote more studying in high school
 - B) are better measures of academic success
 - C) represent a more reliable criterion
 - D) can be easily administered by school personnel
54. If in addition to high school gpa and SAT scores, we also know scores on achievement tests for college applicants, then:
- A) our prediction of college grades will improve substantially
 - B) we would expect a substantial increase in predictive validity
 - *C) our prediction of college grades will increase slightly
 - D) we would find that achievement tests are better than the SAT

55. If we have high school gpa and achievement test scores and then we add SAT scores, our predictive validity will:
- A) increase substantially
 - B) decrease slightly
 - C) decrease substantially
 - *D) stay about the same
56. In the study by Halpin et al. (1981) of the California Achievement Tests as predictors of collegiate gpa, the authors concluded that the best predictor was:
- *A) high school GPA
 - B) the SAT
 - C) the California Achievement Tests
 - D) the ACT
57. In Feingold's (1983) study of the predictive validity of WAIS subtests for college achievement, it was found that:
- A) the SAT was still the best predictor
 - B) such subtests are valid for males but not females
 - C) WAIS subtests were better predictors than high school GPA
 - *D) WAIS subtests can be better predictors than other measures
58. Coaching can affect the validity of a test in at least three ways. Which is NOT one of these?
- A) coaching can increase a person's score above their true level
 - *B) coaching can decrease a person's test anxiety
 - C) coaching can allow a person to do their best rather than their typical
 - D) coaching can affect test performance on stable traits
59. We can probably conclude that coaching for the SAT results in:
- A) substantial improvement on both verbal and math scores
 - B) substantial improvement on verbal scores only
 - *C) a small improvement in math scores
 - D) a decline in verbal scores
60. The SAT was specifically designed to predict
- *A) first year college grades
 - B) academic success in college
 - C) overall GPA in the four years of college
 - D) academic success as defined by graduation
61. Reliability coefficients for the SAT generally range:
- A) from .10 to .50 in most studies
 - B) in the .70 and above range
 - C) from .42 to .83 depending on the type of reliability
 - *D) in the high .80s and low .90s
62. Typically we would expect a correlation of about _____ between SAT scores and first-year college GPA.
- A) .10
 - *B) .40
 - C) .60
 - D) .75
63. When various sources of error are statistically controlled, the average validity of the SAT is about:
- A) .15
 - B) .40
 - *C) .55
 - D) .75

64. Nader presented data to argue that SAT scores were correlated with family income. His data however, is flawed because:
- *A) he computed the correlation using grouped data
 - B) his data was based on small, atypical samples
 - C) there are many ways to measure family income
 - D) the data was obtained from white samples only
65. One of the major limitations of the GRE:
- A) has marginal reliability at best
 - B) does not assess analytical skills
 - *C) does not measure all the qualities related to graduate school success
 - D) scores on the test do not correlate with criteria of success
66. In making admissions decisions, graduate departments typically weigh most heavily:
- *A) undergraduate grades
 - B) letters of recommendation
 - C) GRE scores
 - D) autobiographical statements
67. To do well on the Quantitative portion of the GRE, a candidate should have:
- A) college level courses in algebra and calculus
 - B) considerable work in advanced math classes
 - C) math preparation comparable to 8th grade level
 - *D) typical high school courses in math and algebra
68. The Analytical portion of the GRE attempts to measure such skills as:
- *A) logical and analytical reasoning
 - B) data interpretation and quantitative reasoning
 - C) analogies and analytical comprehension
 - D) knowledge of the arts and humanities
69. The Subject Tests of the GRE
- A) cover broad areas like arts and humanities
 - *B) cover specific areas like Biology and Sociology
 - C) contain the same number of items regardless of field
 - D) are all designed to assess verbal reasoning
70. One feature of the Subject Tests of the GRE:
- A) scores are directly comparable from test to test since they use the same scale
 - B) each test uses a different scaling procedure
 - *C) scores on one test cannot be directly compared to scores on another
 - D) the norms are based on the same group of candidates
71. John scores at the 65th percentile on the Psychology Subject Test and at the 78th percentile on the Sociology Subject Test. We can therefore conclude that:
- A) he should apply to the Sociology department
 - B) he did significantly better on the Sociology test
 - C) the Psychology test must be easier than the Sociology
 - *D) none of these
72. Items for the GRE General Test are written primarily by:
- *A) ETS staff
 - B) university professors
 - C) professional test writers
 - D) hired college students

73. Items for the GRE subject tests are written primarily by:
- A) ETS staff
 - B) university professors
 - *C) experts in that field
 - D) hired college students
74. Studies have shown that individuals who repeat the General test show an average score
- A) decline of about 10 points
 - *B) gain of about 1/4 standard deviation
 - C) score gain of about 50 points
 - D) score gain of about 1 standard deviation
75. In a (an) _____ test, the selection of questions is tailored to an examinee's ability level.
- *A) adaptive
 - B) computer administered
 - C) ipsative
 - D) GRE type
76. In general, the reliability of the GRE Subject Tests is assessed by the _____ procedure.
- A) test-retest
 - *B) Kuder-Richardson
 - C) alternate forms
 - D) Cronbach Alpha
77. Typical validity coefficients for the GRE General Test:
- A) .00 to .20
 - *B) .20 to low .30s
 - C) low .30s to high .50s
 - D) .40s to low .60s
78. To predict graduate first year gpa, the best combination of variables seems to be:
- *A) the three GRE subscores and undergraduate gpa
 - B) GRE Verbal and undergraduate gpa
 - C) GRE Analytical and SAT Total
 - D) letters of recommendation, GRE Total, and biographical statement
79. Which of these variables would we expect to correlate highest with graduate first year gpa in the field of Biology?
- A) undergraduate gpa
 - B) GRE Verbal scores
 - *C) scores on the Biology Subject Test
 - D) GRE Analytical scores
80. One of the major reasons why validity coefficients on the GRE are so low:
- A) the test is psychometrically inadequate
 - B) validity coefficients cannot be higher than reliability
 - C) lack of good criteria
 - *D) restriction of range
81. Marston (1971) compared GRE scores to subsequent professional publications and found:
- A) a modest but significant relationship
 - B) a relationship with GRE Verbal but not with GRE Quantitative
 - *C) a lack of relationship between the two
 - D) that GRE scores were better predictors than GPA
82. Advanced Psychology Test scores
- *A) predict graduate grades but not faculty ratings
 - B) predict completion of graduate degree
 - C) have little relationship to any criteria of success
 - D) seem valid for clinical students but not others

83. Goldberg and Alliger (1992) in their meta analysis of the GRE concluded that we should
- A) throw out the GRE
 - B) only use Advanced Tests
 - C) require the GRE of all applicants
 - *D) focus on the criterion
84. In studies of the validity of the GRE which criterion has been used most often?
- *A) graduate GPA
 - B) undergraduate GPA
 - C) evidence of professional accomplishment
 - D) faculty ratings
85. Some experts regard _____ the best criterion of graduate academic success.
- A) GPA
 - *B) obtaining the degree
 - C) faculty ratings
 - D) number of publications
86. Whitworth and Barrientos (1990) studied Anglo and Hispanic graduate students and the GRE. They found that:
- *A) Anglos scored higher than Hispanics on both GRE and gpa
 - B) Anglos scored higher on GRE but Hispanics scored higher on gpa
 - C) regression equations based on GRE variables predicted gpa well
 - D) the data supported continued use of the GRE
87. The Jones and Thomae-Forgues (1984) review of the MCAT concluded that:
- A) MCAT scores contributed little by way of predictive validity
 - B) the Chemistry and Biology subtests had little validity
 - C) the Physics subtest correlated the most with grades
 - *D) MCAT scores were similar to undergraduate gpa in predictive value
88. In addition to first-year medical school grades, the MCAT is also validated against:
- *A) scores on the National Board of Medical Examiners exams
 - B) professional success as a physician
 - C) faculty ratings of competence as intern or resident
 - D) undergraduate gpa
89. Donnelly et al. (1986) developed a regression equation using MCAT scores and grades in anatomy courses. They found that the regression equation predicted _____ with an accuracy of about _____.
- A) first year medical school grades, 92%
 - B) faculty ratings, 81%
 - C) scores on the NBME-I exams, 15%
 - *D) performance on the Board exams, 92%
90. Based on somewhat remote data, we would estimate the “average true validity” of the MCAT to be about:
- A) .10
 - B) .30
 - *C) .50
 - D) .80
91. Nowacek et al. (1987) studied medical school students and concluded that the pattern of correlations between grades, MCAT scores, and performance on the Board exams reflected:
- A) trait effect
 - B) slope bias
 - *C) method effect
 - D) intercept bias

92. Carline et al. (1983) focused on the last two years of medical school as the criterion and found that:
- A) MCAT scores were excellent predictors of NBME-II exam scores
 - *B) the MCAT was a better predictor of NBME-II than undergraduate science gpa
 - C) undergraduate science gpa was the best predictor of medical school gpa
 - D) medical school gpa and NBME-II scores did not correlate
93. Studies of the MCAT with black medical students suggest that the
- A) MCAT is biased against minority applicants
 - B) MCAT is of very limited value with minority students
 - C) the best predictor of medical school gpa is undergraduate gpa
 - *D) the results are similar to those with white medical students
94. On the MCAT coaching seems to improve scores
- *A) on several subtests
 - B) on all subtests
 - C) for minority students only
 - D) for majority students only
95. The Dental Admission Test Program (DAT) is designed to measure:
- A) the expected prerequisites relevant to the practice of dentistry
 - B) verbal, quantitative, and analytical abilities
 - *C) general academic ability and scientific understanding
 - D) interest and aptitude for being a dentist
96. Scores on the DAT are reported on a scale
- A) similar to the GRE, with mean of 500 and SD = 100
 - B) that uses percentiles
 - *C) that ranges from -1 to +9, with a mean of 4
 - D) of T scores, with mean of 50 and SD = 10
97. In regards to the reliability of the DAT:
- *A) because of speed, test-retest and parallel forms should be reported
 - B) the test-retest reliability is substantially higher than K-R
 - C) there is little evidence to support its reliability
 - D) its internal consistency is consonant with the structure of 4 sections
98. A criticism of the Dental Admission Testing Program:
- *A) may assess academic skills rather than dental skills
 - B) does not predict dental gpa as well as undergraduate gpa
 - C) the test is too short and hence not highly reliable
 - D) the test is too new to fully evaluate
99. Licensing refers to:
- A) a high degree of competency
 - *B) minimal competencies
 - C) membership in a particular organization
 - D) compliance with a local building code
100. The major purpose of licensing exams is to:
- *A) protect the public's welfare and safety
 - B) identify practitioners with outstanding competencies
 - C) predict job success
 - D) assess a person's skills in a specific profession
101. For licensing/certification tests, _____ validity is relatively easy to establish.
- A) criterion
 - *B) content
 - C) construct
 - D) multiple

102. The method of predictive yield is basically:
- A) a logical approach to eliminating slope bias
 - B) a way of analyzing false positives and false negatives
 - C) the result of expert judgment
 - *D) a method to establish a cutoff score
103. The Angoff method of setting a cutoff score involves:
- A) computing a difficulty level of 80%
 - B) the expected chance score
 - C) using a z score of +1 as the dividing line
 - *D) deciding how a barely qualified person would answer
104. A criticism of the various criterion-referenced methods to calculate cutoff score:
- A) result in too lenient of a cutoff score
 - B) use very complex statistical formulas
 - *C) poor interjudge reliability and time consuming
 - D) are substantially intercorrelated with each other

14 Occupational Settings

1. In employment testing most tests are administered to:
 - A) clerical workers
 - *B) middle managers and supervisors
 - C) executives
 - D) assembly line workers
2. The 1968 Testing and Selection Order:
 - A) all job applicants were to be tested
 - B) only three categories of validity were recognized
 - *C) job applicants could not be discriminated against
 - D) government contracts required a job analysis
3. In the Ryan and Sackett (1987) survey of industrial psychologists, interviews were the most common assessment practice followed by:
 - *A) personal history forms
 - B) ability tests
 - C) projective tests
 - D) simulation exercises
4. In the Ryan and Sackett (1987) survey of industrial psychologists, the most frequently used test was the:
 - A) WAIS
 - B) MMPI
 - C) In-basket
 - *D) Watson-Glaser Critical Thinking Appraisal
5. In the Ryan and Sackett (1987) survey of industrial psychologists, the least valid procedure was judged to be:
 - A) simulations
 - B) ability tests
 - *C) projective techniques
 - D) personal history forms
6. Many of these tests involve apparatus rather than paper-and-pencil:
 - A) intellectual abilities
 - B) motivation
 - C) biodata
 - *D) sensory and psychomotor abilities
7. In the Nathan and Alexander (1988) meta analysis, the least valid method of measuring job performance was:
 - *A) quality of production
 - B) supervisor ratings
 - C) work samples
 - D) quantity of production

8. Ghiselli and Haire (1960) suggested that test criteria are “dynamic.” This means that:
 - A) criteria are highly unreliable and should not be used
 - *B) the criterion may be different at different times
 - C) multiple criteria should be used
 - D) test results will not correlate significantly with any criteria
9. The Ghiselli and Haire (1960) study of taxicab drivers was discussed to illustrate the idea that:
 - A) tests that are valid will correlate significantly with any criteria
 - B) in a work setting tests typically do not correlate with criteria
 - *C) criteria need to be selected and specified carefully
 - D) because criteria are “dynamic” it is useless to measure them
10. Weitz’s (1950) study of preferences about a cooking range illustrates the point that:
 - A) interview methods are quite robust and produce equal results
 - *B) different methods of asking the same question produce different results
 - C) a test can predict criterion X but not criterion Y
 - D) interitem consistency seems to be the best type of reliability
11. Hunter and Hunter (1984) did a meta analysis and reported that cognitive tests have a mean validity of about _____ in predicting training success.
 - A) .15
 - B) .35
 - *C) .55
 - D) .85
12. Hunter and Hunter (1984) on the basis of their meta analysis concluded that the validity of psychomotor tests:
 - A) is too low to be used in any practical way
 - B) is exceptionally high and shows these tests to be useful
 - C) can be quite high if administered by well trained examiners
 - *D) can under some circumstances be very low
13. Overall, the two most valid predictors of job performance are:
 - *A) cognitive ability tests and biodata forms
 - B) tests of personality and of psychomotor skills
 - C) biodata forms and simulation exercises
 - D) interviews and simulation exercises
14. In the Ree and Earles (1991) study of airmen, the most valid predictor of technical school grades was:
 - A) spatial intelligence
 - *B) general cognitive ability
 - C) psychomotor skills
 - D) instructors’ ratings
15. Differential validity, when it occurs, is typically:
 - A) a reflection that the test is biased
 - B) a result of using invalid criteria
 - C) based on real differences between samples
 - *D) an artifact of small sample size
16. In applying the multitrait-multimethod approach to performance appraisal:
 - A) multitrait is replaced by multiraters
 - B) multitrait is replaced by monotrait
 - *C) multimethod is replaced by multiraters
 - D) multimethod is replaced by multicriteria

Occupational Settings

17. In the Lawler (1967) study of the multitrait-multimethod approach, the validity coefficients were the _____ coefficients.
 - A) heterotrait-monorater
 - B) heterotrait-heterorater
 - C) monotrait-monorater
 - *D) monotrait-heterorater
18. In the Lawler (1967) study of the multitrait-multimethod approach, the one variable that did not show validity was:
 - *A) self ratings
 - B) peer ratings
 - C) quality of job performance
 - D) ability to perform the job
19. The Harris and Schaubroeck (1988) meta analysis of ratings found that:
 - A) peer and supervisor ratings did not correlate with each other
 - *B) self ratings showed low correlations with peer and supervisor ratings
 - C) peer ratings were least valid of all ratings
 - D) self ratings were the most valid
20. A review of the literature on the halo error concludes that:
 - A) the halo effect is a rather common problem in ratings
 - B) valid and invalid components can be identified reliably
 - *C) the halo effect does not necessarily detract from validity
 - D) typically the halo effect lowers the reliability of ratings
21. In regards to rater bias:
 - A) the literature indicates that can result in higher validity
 - B) we can use the multitrait-multimethod approach to control it
 - *C) there are several methods to counter this
 - D) it is a major problem only with ratings of cognitive competence
22. A leniency error means that the rater
 - A) is lenient and rates everyone as above average
 - *B) tends to use middle or average ratings
 - C) is highly biased and his/her ratings reflect this
 - D) is not reliable in his/her ratings
23. In the _____ method, the rater selects the best and the poorest persons, then the next best and next poorest, etc.
 - A) forced distribution
 - B) paired comparison
 - C) high-low
 - *D) alternation ranking
24. The usefulness of the paired comparison method is limited by:
 - *A) the number of persons to be rated
 - B) the number of raters required
 - C) the rectangular distribution formed by ranks
 - D) such errors as the leniency error
25. One of the limitations of supervisors' ratings:
 - A) most supervisors have little contact with their employees
 - *B) supervisors may not have observed the behavior to be rated
 - C) the reliability of such ratings is very poor
 - D) such ratings do not correlate with peers' ratings

26. Behavioral anchors on a rating scale
 - A) are actually generated by the person doing the ratings
 - B) vary from target to target
 - *C) provide a standard framework to evaluate performance
 - D) can only be used with personality dimensions
27. Self ratings are typically $\frac{1}{2}$ SD higher than supervisory ratings. This means that:
 - *A) subordinates tend to be more lenient than supervisors in self evaluations
 - B) supervisors tend to be more lenient in rating workers
 - C) when subordinates rate their supervisors, they are more lenient
 - D) self ratings lack validity
28. Some evidence suggests that in Taiwan, Chinese
 - A) supervisors rated employees less harshly than Anglo supervisors
 - B) employees rate their peers as excellent workers
 - C) supervisors base their ratings primarily on group consensus
 - *D) employees rated their job performance less favorably than did their supervisors
29. The attraction-selection-attrition framework refers to the concept that:
 - *A) different types of people make up different types of organizations
 - B) whether two individuals are attracted to each other depends on personality
 - C) personality is the best predictor of job performance
 - D) individuals selected for a job are selected because of personality traits
30. In the Barrick and Mount (1991) analysis of job performance criteria for five occupational groups:
 - A) extraversion was the most valid predictor for police
 - B) both conscientiousness and extraversion were of limited validity
 - *C) conscientiousness showed consistent relations with job performance
 - D) no distinction was made between job proficiency and training proficiency
31. In general, most correlations between well designed personality inventories and job effectiveness criteria fall in the _____ range.
 - A) .00 to .20
 - *B) .40 to .60
 - C) .50 to .70
 - D) .60 to .80
32. Hogan (1990) on the basis of a meta analysis of the Big Five model, concluded that:
 - A) emotional stability is correlated with supervisors' ratings
 - B) openness to experience is correlated with leadership
 - C) likability is correlated with academic performance
 - *D) extraversion is correlated with performance in sales and management
33. Hogan (1990) on the basis of a meta analysis of the Big Five model concluded that _____ was related to upward mobility and leadership status.
 - A) extraversion
 - *B) emotional stability
 - C) agreeableness
 - D) dependability
34. In regards to the Big Five model and the area of personnel selection, we might conclude that:
 - A) the model has excellent validity across studies
 - B) the 16 PF provides a better factor analytic framework
 - *C) the model is useful though limited
 - D) the model has little validity to support it

35. In the Kelley et al. (1994) study of the MMPI of nuclear power workers, the results indicated that:
- *A) test-retest differences were rather small
 - B) the typical test-retest reliability was .35
 - C) upon retest, test profiles tended to be more abnormal
 - D) MMPI test scores were excellent predictors of depression among workers
36. The Chinese Value Survey measures
- A) the degree to which a person follows Confucian teachings
 - B) the degree of acculturation in Chinese immigrants
 - C) the values that distinguish poor from good managers in Taiwan
 - *D) values of importance in eastern countries like China
37. In the study by Ralston et al. (1992) where the Chinese Value Survey was administered to managers from three countries, the results showed:
- A) that managers in Hong Kong had the highest values
 - *B) the U.S. managers scored highest on human heartedness
 - C) the Confucian Work Dynamism scale was not valid
 - D) significant differences on Moral Discipline
38. In regards to biodata, one major source of concern is:
- *A) their susceptibility to faking
 - B) they are not readily available to the public
 - C) their test-retest reliability tends to be low
 - D) they are considered by many an invasion of privacy
39. Biodata items typically focus on either a particular domain (like job-related skills) or
- A) cognitive skills
 - B) personality variables
 - *C) previous life experiences
 - D) future aspirations
40. A typical biodata item:
- A) Do others consider you a friendly person?
 - B) What is the square root of 64?
 - C) How important is it for you to be loyal?
 - *D) How many years have you worked in sales?
41. In developing a biodata form we might do some preliminary statistical analyses on tentative items to determine:
- A) split-half reliability
 - *B) whether they show skewed response distributions
 - C) their tertiary validity
 - D) whether the items correlate with simulation exercises
42. In the empirical keying approach we would administer a pool of potential biodata items
- A) wait a period of time, and then retest all subjects
 - B) and retain items based on internal consistency
 - *C) and develop a scoring procedure based on some criterion
 - D) and retain items with high factor loadings
43. In the empirical keying procedure, a crucial aspect is the:
- A) type of factor analysis carried out
 - B) age of criterion group members
 - C) use of the indirect approach of item construction
 - *D) reliability, validity, and utility of the criterion

44. In a(n) _____ biodata questionnaire, different weights are given to different responses.
- *A) weighted
 - B) rational
 - C) empirical
 - D) multivariate
45. Mumford and Owens (1987) reviewed the empirical keying approach and concluded that:
- A) the approach is quite valid but shows significant cultural differences
 - B) factorial scales represent a much better approach
 - *C) such biodata forms show considerable cross-cultural validity
 - D) the approach works well for personality tests but not biodata forms
46. A major criticism of the empirical keying approach is that:
- A) the empirical validity of such scales is quite low
 - *B) there may be a lack of construct validity
 - C) the required job analysis is too time consuming
 - D) any cultural differences are disregarded
47. The direct and indirect strategies of developing biodata forms
- A) are identical, but the resulting scoring procedures differ
 - B) differ in their use of factor analysis to assess construct validity
 - C) are basically identical but differ in their focus on internal consistency
 - *D) differ in that the indirect approach focuses on constructs that underlie job performance
48. Studies that compare rational versus empirical scaling techniques suggest that:
- *A) empirical scales have higher initial criterion related validity
 - B) empirical scales are substantially more valid in the long run
 - C) rational scales tend to be less reliable than empirical scales
 - D) both approaches are equally valid
49. Items for biodata scales developed by the rational method are retained based on:
- *A) their internal consistency
 - B) the author's judgment
 - C) consensus by a panel of experts
 - D) high test-retest reliability
50. Items for biodata scales developed by factor analysis are retained if they load at least _____ on a particular dimension.
- A) .20
 - *B) .30
 - C) .50
 - D) .80
51. Mumford and Owens (1987) suggest that factorial biodata scales may not be as effective as empirically keyed scales, but
- A) are easier to translate using the back translation method
 - B) are easier to construct
 - *C) may have greater stability and generality
 - D) yield more understandable results
52. The option-keying scoring procedure involves
- A) using Likert weights (5 to 1) for each response option
 - B) factor analyzing the response matrix
 - C) using true-false rather than multiple choice items
 - *D) analyzing each response alternative separately

Occupational Settings

53. In the Mitchell and Klimoski (1982) study of biodata of real estate salespersons, the results showed that:
- *A) empirical scoring was superior but showed shrinkage upon cross-validation
 - B) rational scoring was superior but showed shrinkage upon cross-validation
 - C) rational scoring was better than the empirical in both original and cross-validation samples
 - D) empirical scoring was better but only in the original sample
54. Mumford and Owens (1987) suggested that faking on biodata questionnaires
- A) is difficult to control
 - B) has been shown beyond a doubt
 - C) significantly lowers their validity
 - *D) can be minimized
55. Reliability of biodata items can be increased by:
- A) making items simple and brief
 - B) giving response options on a numerical continuum
 - C) providing an “escape” response option
 - *D) all of the above
56. An “escape” response option might take the form:
- *A) none of the above
 - B) my salary was less than \$10 an hour
 - C) I do not wish to answer this item
 - D) I was fired from my last job because. . .
57. Well developed biodata questionnaires typically show:
- A) substantial internal consistency
 - B) low test-retest reliability
 - *C) low item intercorrelations
 - C) high factor validity
58. In general, the literature indicates that biodata forms
- A) correlate on average .80 or above with job productivity
 - *B) are valid in a wide variety of settings and samples
 - C) show significant black-white differences
 - D) show low correlations between self-report and employer report
59. In the Scott and Johnson (1967) study of canning factory workers, the regression equation gave the heaviest weight to the variable of:
- A) type of residence
 - B) number of children
 - C) years on last job
 - *D) gender
60. In the Scott and Johnson (1967) study of canning factory workers, the results showed that _____ stayed longer on the job.
- A) skilled male workers who lived with parents
 - B) skilled female workers with great family responsibilities
 - *C) unskilled females living close to work
 - D) workers who lived at some distance from the factory
61. Assessment centers typically involve:
- *A) multiple assessors and rating forms
 - B) computerized tests administered individually
 - C) a physical office where evaluations are carried out
 - D) tests of intelligence and of achievement

62. Meta-analytic reviews of assessment centers support their:
- A) internal consistency
 - *B) predictive validity
 - C) standardization
 - D) construct validity
63. Studies of why assessment centers work suggest that:
- A) employees who are assessed are highly pre-selected
 - B) both convergent and discriminant validities are quite high
 - C) the obtained ratings are highly biased
 - *D) a global dimension like intelligence or competence may be central
64. In-basket techniques are usually scored for
- A) originality and productivity
 - B) efficiency and competency
 - *C) content and style
 - D) industry and leadership
65. Of the various types of reliability, the in-basket technique shows the highest results on _____ reliability, and the lowest results on _____ reliability.
- *A) interrater; alternate form
 - B) split-half; internal consistency
 - C) alternate form; test-retest
 - D) internal consistency; split-half
66. Schippmann et al. (1990) conclude that the validity of the in-basket technique
- A) is quite solid despite some negative evidence
 - B) can be quite high if experienced assessors are used
 - C) has not been adequately explored
 - *D) is based more on belief than evidence
67. In the study by Brannick et al. (1989) with two alternate forms of the in-basket exercise used with business students:
- *A) scores for the two forms on the same dimensions correlated low
 - B) interjudge reliability was very low for all dimensions
 - C) internal reliability coefficients were uniformly high
 - D) only one of the two forms proved to be valid
68. A good way to communicate predictive validity results is through a(n):
- A) work sample
 - *B) expectancy table
 - C) simulation exercise
 - D) concept of banding
69. The concept of banding involves:
- A) treating all test scores as if they were identical
 - B) using the median to define above and below average
 - *C) ignoring numerical differences in scores smaller than the band value
 - D) having subjects act as a team and scoring their performance as a group
70. Synthetic validity:
- A) an alternative model to face validity
 - B) applicable only to assessment centers
 - C) assesses the validity of each item in a test
 - *D) validating tests against specific job elements
71. Time urgency:
- *A) is most likely a multidimensional construct
 - B) shows low coefficient alpha reliability
 - C) is a major dimension of the in-basket technique
 - D) is not related to aspects like competitiveness

Occupational Settings

72. In the Cascio (1975) study of turnover in insurance clerical employees, it was found that:
- A) biodata items did not predict long vs. short tenure
 - *B) a 10 item biodata score was quite valid
 - C) the biodata did not work well with minority employees
 - D) there were significant gender differences in turnover
73. In the area of occupational choice, biodata questionnaires:
- A) are most particularly useful since they do not measure career interests
 - B) are of limited use since their reliability is low
 - *C) correlate significantly with post college job choice
 - D) are useful to assess creativity but not vocational choice
74. The Navy Basic Test Battery is validated primarily against
- A) scores on the General Classification Test
 - B) performance aboard ship in military service
 - C) career achievements like rank and pay
 - *D) school grades for trainees at the end of their training
75. The DOT Estimation Task was designed to assess
- A) visual-spatial abilities of Air Force pilots
 - B) creativity versus conservativeness
 - *C) compulsivity versus decisiveness
 - D) visual acuity in night-time conditions
76. The Imwald Personality Inventory was designed to assess the:
- *A) psychological and emotional fitness of police recruits
 - B) mental ability of potential police officers
 - C) degree to which a police officer follows the law
 - D) career interests people may have in regards to police work
77. In the Baehr et al. (1971) study using a test battery to select police patrolmen, racial differences on the predictor variables
- A) were not found for any of the four areas
 - *B) were found for three of the four broad domains
 - C) were found only for the area of mental ability
 - D) were found for both aptitude and motivation
78. In the Baehr et al. (1971) study of police officers the investigators found that:
- A) white and black officers differed on tenure and absenteeism
 - B) annual performance ratings were lower for black officers
 - C) there were no significant racial differences on job performance
 - *D) black patrolmen made more arrests and had more disciplinary demerits
79. The Baehr et al. (1971) study of police officers concluded that:
- A) good police officers have a social rather than a work orientation
 - B) the test battery they used was not valid for black officers
 - *C) the successful patrolman tends to be cooperative in interpersonal situations
 - D) there is no need to compute separate regression equations for blacks and for whites
80. The Job Components Inventory is designed to:
- A) help college students make better career choices
 - B) assess engineering aptitude
 - *C) assist young people in their vocational preparation and training
 - D) predict whether a person will enter a clerical position

81. In developing the Supervisory Practices Test, items were retained if they discriminated between supervisors and non-supervisors, and if:
- A) their test-retest reliability was at least .70
 - B) a panel of experts agreed on the utility of the item
 - C) they had a minimal loading of .30 on a factor
 - *D) the distractors were endorsed by at least 15%
82. The authors of the Supervisory Practices Test agreed that the test-retest reliability was a low estimate because of restriction of range. However:
- A) test-retest is not the appropriate procedure to use
 - *B) raw scores seem to be quite varied
 - C) restriction of range affects validity not reliability
 - D) they should have computed an alpha coefficient
83. What was unique about the General Aptitude Test Battery at its inception?
- *A) it was a comprehensive battery
 - B) it assessed 12 basic aptitudes
 - C) five forms were available for various purposes
 - D) none of the subtests had time limits
84. Both the "Place and Turn" and "Assemble and Disassemble" subtests of the GATB:
- *A) are performance subtests
 - B) have poor equivalent form reliability
 - C) are composed of very few items
 - D) do not have time limits
85. In terms of temporal stability of the GATB:
- *A) the highest test-retest coefficients are found on the cognitive aptitudes
 - B) typical equivalent form reliability coefficients are in the .80s
 - C) the GATB is a speeded test so internal consistency is not appropriate
 - D) highest for the psychomotor aptitudes
86. Much of the validity information on the GATB can be interpreted as _____ validity.
- A) factorial
 - *B) construct
 - C) predictive
 - D) convergent
87. What is a major criticism of the GATB?
- A) little information about reliability and validity
 - B) developed in England and may not be useful in the U.S.
 - *C) very complicated statistical scoring procedure
 - D) cannot be administered at a single setting
88. The How Supervise? test items were initially selected from a pool of potential items by
- *A) expert judgment
 - B) high test-retest reliability
 - C) significant correlations with criteria
 - D) discriminant analysis of supervisors and controls
89. It appears that the How Supervise? test might actually be a measure of:
- A) managerial success
 - B) test sophistication
 - *C) intelligence
 - D) motivation

Occupational Settings

90. The two motivational aspects related to high quality managerial success are:
- A) job proficiency and self-discipline
 - B) responsibility and dependability
 - *C) need for advancement and high personal standards
 - D) supervisory practices and supervisory opinions
91. High scorers on the Work Orientation scale would typically be described as:
- A) highly intelligent but not creative
 - B) both very intelligent and very creative
 - C) socially oriented but immature
 - *D) following the "protestant ethic" of work
92. The Wonderlic Personnel Test is designed to assess:
- A) how well organized and dependable the person is
 - *B) general mental ability
 - C) motivation for advancement
 - D) leadership ability
93. In regards to converting test scores to percentiles on the Wonderlic Personnel Test, the author recommends that:
- A) this not be done; only raw scores are to be used
 - B) local norms be established
 - *C) separate norms by race and ethnic group be used
 - D) national norms rather than racial norms be used
94. Since older individuals tend to do less well on the Wonderlic Personnel Test, the author recommends that:
- A) the test not be administered to persons older than 40
 - B) specific age norms be used to account for age
 - *C) a specific number of points be added to reflect a person's age
 - D) older individuals be given a more liberal time limit
95. Integrity tests are typically "proprietary" tests. This means that:
- A) they are published commercially but available free of charge
 - B) they are carefully worded and use proper English
 - C) the copyright belongs to a professional organization
 - *D) they are closely controlled and used under strict guidance
96. A characteristic of current integrity tests:
- A) only one or two are used in industry
 - *B) little psychometric information available in scientific journals
 - C) have rather low reliability
 - D) their predictive validity is not supported
97. The base rate for a behavior like employee theft is usually:
- *A) less than 5%
 - B) about 20%
 - C) about 50%
 - D) over 50%
98. Scoring keys for integrity tests are usually:
- A) available only to qualified buyers
 - B) found in such reference works as the MMY
 - C) readily available through publisher's catalogs
 - *D) closely guarded by the publisher

99. Overt integrity tests usually have two components:
- A) job satisfaction and the prediction of burnout
 - B) deviant behavior and irresponsibility
 - *C) attitudes towards theft and admission of illegal behavior
 - D) past behavior and future behavior
100. Integrity tests _____ with polygraph results.
- *A) correlate significantly
 - B) do not correlate significantly
 - C) have low but significant correlations
 - D) are usually administered
101. The correlation between integrity test scores and future behavior is typically:
- A) high, especially when measured by polygraph studies
 - B) low because of shrinkage reduction
 - *C) low because of the low base rate
 - D) high, especially when samples of convicts are used
102. One of the major “problems” of integrity tests is that:
- A) there are too many of them to come to any conclusion
 - *B) the variance on the criterion of theft is highly restricted
 - C) they typically do not discriminate between convicts and the general public
 - D) they do not correlate significantly with polygraph results
103. In regards to adverse impact of integrity tests, we can conclude that:
- *A) there are no significant gender and/or race differences
 - B) minority members tend to score substantially lower
 - C) such tests are best not used with minority members
 - D) the evidence suggests such tests are biased against blacks
104. In the Collins and Schmidt (1993) study of white-collar crimes, there were four major variables that discriminated between criminals and non-criminals. The authors felt that these four variables had a common theme of:
- A) honesty and integrity
 - B) intelligence
 - *C) social conscientiousness
 - D) criminality

15 Clinical and Forensic Settings

1. The Halstead-Reitan battery takes about _____ to administer.
A) 15 minutes
B) two hours
*C) five hours
D) eight hours
2. Most of the subtests used in the Halstead-Reitan
A) were developed specifically for this battery
*B) were actually borrowed from other procedures
C) are very sound psychometrically
D) are easy to administer and score
3. The validation of the Halstead-Reitan has primarily focused on _____ validity.
A) construct
B) content
C) predictive
*D) concurrent
4. The Rorschach Inkblot Technique is an example of a(n) _____ technique.
*A) associative
B) construction
C) ordering
D) completion
5. The reliability of qualitative data is often assessed by:
A) test-retest
*B) accuracy of protocol matches
C) interrater reliability
D) internal consistency methods
6. Both methods of establishing the reliability of qualitative data have the limitation that:
A) experts are required to carry out the procedure
B) reliability and validity are confounded
C) test reliability requires a test-retest design
*D) test reliability is confounded with the judgment process
7. In presenting the Rorschach Inkblots to a client, the sequence of presentation:
*A) is always the same
B) is up to the administrator
C) changes with the client's gender
D) does not much matter

8. The purpose of the inquiry phase on the Rorschach is to obtain information
 - A) about the client and his/her background
 - *B) so the responses can be scored
 - C) that determines which inkblots will be administered
 - D) as to how normal or psychotic the client is
9. Form, color, and shading aspects of the Rorschach are known as:
 - A) content variables
 - *B) determinants
 - C) location
 - D) testing the limits
10. A perceptual-nomothetic approach to the Rorschach would focus on:
 - *A) how subjects in general perceive the inkblots
 - B) what subjects in general perceive in the inkblots
 - C) how a specific subject perceives the inkblots
 - D) what a specific subject perceives in the inkblots
11. In the Albert et al. (1980) study of faking on the Rorschach, most of the faked protocols were identified as:
 - A) faked
 - B) normal
 - C) neurotic
 - *D) psychotic
12. The fact that there are only 10 inkblots in the Rorschach means that
 - A) validity cannot be established
 - *B) reliability is inherently limited
 - C) only alternate form reliability can be computed
 - D) reliability cannot be determined
13. Parker's (1983) meta analysis of Rorschach studies indicated that reliabilities of about _____ can be expected in well designed studies.
 - A) .50
 - B) .60
 - C) .70
 - *D) .80
14. Percentage of agreement is a poor measure of reliability because:
 - *A) it does not take into account chance agreement
 - B) it requires too many experts to calculate
 - C) it is confounded by validity results
 - D) experts typically do not agree
15. In the Holtzman Inkblot Technique, the subject
 - A) gives five responses per card
 - *B) gives one response per card
 - C) takes both sets of cards
 - D) responds only to those inkblots that are meaningful
16. In general, the reliability of the Holtzman is:
 - A) adequate with most coefficients in the .60-.70 range
 - B) rather poor with typical results below .50
 - *C) high with typical coefficients in the .80s and .90s
 - D) low for intra-scorer reliability but adequate for test-retest

17. Validity studies of the Holtzman typically:
 - A) report negative findings despite high reliability
 - *B) compare diagnostic groups or relate variables to personality test scores
 - C) compare HIT variables with Rorschach variables
 - D) find that results are a function of the scoring method used
18. The TAT consists of
 - A) two sets of 45 inkblots
 - B) a series of stories chosen for their emotional impact
 - C) a series of pictures depicting the adventures of Blacky
 - *D) a set of 31 pictures, most of which are ambiguous
19. In administering the TAT most clinicians
 - A) have a favorite subset of cards that they administer to every client
 - B) administer the entire set of pictures
 - *C) select a subset of 6 to 10 cards related to the specific client
 - D) ask the client which pictures are best and least liked
20. The “pull” of the card refers to:
 - *A) common themes elicited by a particular card
 - B) whether a subject touches a card and thus shows compliance
 - C) the influence of a picture on subsequent stories
 - D) how often that card is selected for administration
21. The theory behind the TAT focuses on:
 - A) achievement and failure
 - B) castration anxiety
 - *C) needs and presses
 - D) self esteem and competence
22. In regards to the hero of TAT stories, it is assumed that
 - A) the hero represents what Freud called the Id
 - *B) the client identifies psychologically with the hero
 - C) the hero is always a masculine figure
 - D) in each story there are several heroes
23. In a clinical setting, the analysis of the TAT
 - A) follows a strict scoring procedure based on 36 needs
 - B) assesses such aspects as determinants and form quality
 - C) requires a content analysis of both needs and presses
 - *D) usually reflects an impressionistic, holistic approach
24. In regards to the reliability of the TAT we can conclude that:
 - *A) standard methods of determining reliability are not fully applicable
 - B) test-retest is unusually high for a projective technique
 - C) both split-half and parallel form reliabilities are adequate
 - D) the TAT is a research tool not a psychological test
25. In the McClelland and Boyatzis (1982) study of TAT protocols of business managers, they found that:
 - A) need for achievement was significantly related to engineering competence
 - *B) a leadership motive pattern was associated with managerial success
 - C) none of the TAT variables were related to managerial competence
 - D) a need for power correlated with success for all managers
26. Lundy (1988) argued that divergent views and findings on the TAT may in fact reflect:
 - *A) the way the test is administered
 - B) which of the scoring methods (like Exner’s) is used
 - C) which specific cards are selected for administration
 - D) the theoretical presuppositions of the examiner

27. The Washington University Sentence Completion Test attempts to measure:
 - A) need achievement
 - B) adjustment
 - *C) ego development
 - D) interpersonal conflict
28. In the Rotter Incomplete Sentences Blank, the first step in scoring is to:
 - A) assign each response a score of 0 to 6 on adjustment
 - *B) categorize responses as conflict, positive, etc.
 - C) ask the respondent which responses are most accurate
 - D) assess the inter-scorer reliability
29. In general we can conclude that the Rotter has:
 - *A) good reliability but modest validity
 - B) poor reliability and poor validity
 - C) poor reliability but excellent validity
 - D) both excellent reliability and validity
30. The norms on the Rotter:
 - A) are based on several samples of psychiatric patients and of college students
 - *B) are based on a college sample and no longer current
 - C) represent a nationwide sample selected as to census characteristics
 - D) are quite current and varied
31. A test that uses cartoon-like drawings to depict everyday situations:
 - A) the Holtzman Inkblot Technique
 - B) the Thematic Apperception Test
 - C) the Everyday Adjustment Drawings
 - *D) the Rosenzweig Picture-Frustration Study
32. Responses on the Rosenzweig are scored for:
 - A) overall adjustment
 - B) needs and presses
 - *C) direction and type of aggression
 - D) degree of frustration tolerated
33. When the Draw-A-Man is used as a measure of intelligence, using modern scoring systems like Naglieri's, the results indicate:
 - *A) low but generally satisfactory reliability
 - B) higher reliability for each of the drawings than for total score
 - C) poor interrater reliability but acceptable test-retest
 - D) substantially high reliability with typical coefficients in the low .90s
34. The Bender-Gestalt is basically a measure of:
 - A) personality
 - *B) neuropsychological functioning
 - C) drawing ability
 - D) haptic intelligence
35. On the Bender-Gestalt the subject is instructed to
 - A) tell what each design could be
 - B) reproduce each design from memory
 - *C) copy each of nine geometric designs
 - D) select which designs are liked and which disliked
36. The Koppitz system of scoring the Bender-Gestalt
 - A) looks at some 106 scorable features
 - B) has very low interscorer reliability
 - C) assesses the degree of brain damage in an adult
 - *D) involves a set of pathognomonic signs

37. in general, most scoring systems for the Bender-Gestalt have
- A) poor interscorer reliability but high internal consistency
 - B) adequate interscorer reliability but poor test-retest reliability
 - *C) adequate test-retest reliability and high interscorer reliability
 - D) high internal consistency and high test-retest reliability
38. In the review by Tolor and Brannigan (1980) of the Bender-Gestalt, the authors concluded that the test
- A) was not valid as a measure of neuropsychological functioning
 - B) was not valid either to measure personality or neuropsychological functioning
 - C) was basically a measure of general intellectual competence
 - *D) was quite valid as a measure of neuropsychological functioning
39. In the study by Goldberg (1959) where Bender-Gestalt protocols were assessed as to organicity:
- A) degree of confidence was highly related to diagnostic accuracy
 - *B) both psychology trainees and secretaries were more correct than psychologists
 - C) all groups did better than formal scoring
 - D) impressionistic formulations were better than psychometric scoring
40. A conclusion from the Goldberg (1959) study of Bender-Gestalt protocols:
- *A) psychometric analysis is better than impressionistic formulations
 - B) secretaries can assess maladaptive behavior better than psychologists
 - C) the Pascal-Suttell method is better than the Koppitz method
 - D) the Bender-Gestalt has poor reliability and validity
41. Meehl's (1954) review of studies relevant to clinical versus statistical prediction, found that:
- A) most of the studies found statistical prediction to be more accurate
 - B) statistical and clinical predictions were equally accurate
 - C) effective therapists preferred the clinical method
 - *D) no studies found the clinical method to be superior
42. The AB scale on the SVIB was developed by comparing
- *A) therapists successful with schizophrenic patients with non-successful
 - B) high scorers on the Lawyer scale with low scorers
 - C) schizophrenics and neurotics
 - D) clients who are good at problem solving versus poor
43. The term "triage" refers to:
- A) evaluating a therapeutic program for alcohol abuse
 - *B) assignment of patients to specific treatments
 - C) a type of assessment instrument used with alcoholics
 - D) the degree of acceptance that one has a drinking problem
44. The "direct scale" and "indirect scale" strategies to develop scales to assess alcoholism differ as to:
- A) how the initial pool of items is constructed
 - B) the degree of test-retest reliability
 - *C) the relationship of the items to drinking behavior
 - D) objective versus projective items
45. The McAndrew scale of alcoholism was developed by the indirect scale strategy. Which of these items would therefore NOT be included:
- *A) I have used alcohol excessively
 - B) I enjoy the company of others
 - C) In high school I was often in trouble
 - D) My hands shake a lot

- 46. The items on the Michigan Alcoholism Screening Test:
 - A) are all indirectly related to alcohol abuse
 - *B) have high face validity
 - C) were selected because of construct validity
 - D) are based on the notion of an alcoholic personality
- 47. In regards to norms on tests of alcoholism:
 - A) these typically include separate samples of adolescents and women
 - B) usually available based on U.S. census data
 - *C) separate norms for women and minority patients are rarely available
 - D) not particularly useful since alcoholism is not “normal”
- 48. Bulimia involves
 - A) extreme restriction of food intake
 - *B) bouts of binge eating followed by vomiting and/or laxative use
 - C) an extreme phobia that food can be harmful
 - D) a strong dislike for categories of food, like vegetables
- 49. The Eating Disorder Inventory-2
 - A) requires an experienced clinician to administer
 - B) does not consider bulimia
 - *C) is appropriate for both males and females
 - D) yields one total score
- 50. Much of the validity of the EDI-2 focuses on _____ validity.
 - A) face
 - *B) criterion
 - C) factorial
 - D) construct
- 51. In regards to the factorial validity of the EDI-2, we can conclude that:
 - A) the factor structure matches the original eight scales
 - B) there are three rather than eight major factors
 - *C) additional evidence needs to be obtained
 - D) there seems to be two major factors, one of bulimia and one of anorexia
- 52. A personality type marked by peculiar ideas and appearance, as well as deficits in relating to others:
 - A) obsessive compulsive
 - B) type A
 - C) interoceptive
 - *D) schizotypal
- 53. In developing the schizotypal self report measure, Raine (1991) administered the pool of items to:
 - *A) a sample of college students
 - B) schizophrenics and neurotics
 - C) psychiatric patients with a schizotypal diagnosis
 - D) college students identified as schizotypal
- 54. The concept of “hardiness” refers to:
 - A) individuals who can survive in extreme environments
 - *B) resistance to the effects of stressful life events
 - C) a constellation of personality traits related to schizophrenia
 - D) people who have strong appetites

55. Three characteristics seem central to the concept of hardiness; which one does NOT belong?
- A) sense of personal control
 - *B) avoidance of stressful relationships
 - C) commitment to work and activities
 - D) sees life as a challenge
56. Stimulus oriented theories of stress define stress as:
- A) the response of the individual to specific stimuli
 - B) the interaction between event and a person's perception
 - C) subjective beliefs about the world
 - *D) present in selected aspects of the environment
57. The items for the Schedule of Recent Experience were originally generated
- A) by medical experts
 - B) from interviews with psychiatric patients
 - C) by the test authors
 - *D) from the charts of medical patients
58. The relationship between major life changes and health outcomes is:
- A) a substantial one, supported by much evidence
 - *B) modest at best
 - C) a highly controversial issue
 - D) not supported by the research literature
59. The weights assigned to each item on the Schedule of Recent Experience
- *A) reflected the judgment of a sample of subjects
 - B) are empirically related to the severity of illness
 - C) were calculated by the test authors
 - D) were anchored at 100 for death and 10 for marriage
60. In regards to the scoring procedure on the Schedule of Recent Experience
- *A) weighted scores and unitary weights correlate substantially
 - B) weighted scores are superior to unitary weights
 - C) direct magnitude estimation is the most valid method
 - D) unitary weights cannot be used because of the nature of the items
61. How stable are life change scores over time?
- *A) there is stability but only for normal subjects
 - B) there is stability for both psychiatric patients and normal controls
 - C) there is little stability for most groups studied
 - D) quite stable for U.S. subjects but unstable in other cultures
62. The Holmes & Rahe scale has been criticized as lacking some categories of life events. This is a criticism of its _____ validity.
- *A) content
 - B) construct
 - C) criterion
 - D) factor
63. If you were to develop your own checklist of stressful life events, based on the research literature which item would you be most likely to include:
- A) falling down the stairs
 - B) going on vacation
 - *C) having financial difficulties
 - D) sneezing and having a fever
64. The items on the Hassles scale were generated by
- *A) the research staff
 - B) reading medical patients' charts
 - C) interviewing psychiatric patients
 - D) asking college students to keep diaries

65. Which of these is NOT one of the major purposes of health status measures:
- A) assess the effects of clinical interventions
 - B) examine changes in the health care delivery system
 - *C) identify individuals at risk for mental illness
 - D) examine the health of general populations
66. The Sickness Impact Profile is:
- A) designed for adolescents who have social difficulties
 - B) primarily for use with psychiatric patients
 - *C) aimed primarily at elderly medical patients
 - D) an MMPI-like instrument for use with adults
67. One of the administrative problems of the Sickness Impact Profile is that
- *A) handscoring is awkward at best
 - B) the interviewer form is less reliable
 - C) computer scoring is not available
 - D) its manual presents too much validity information
68. The McGill Pain Questionnaire is designed to:
- A) identify individuals who can withstand severe pain
 - *B) provide a quantitative profile of clinical pain
 - C) diagnose individuals who are complainers
 - D) predict which patients suffer from chronic pain syndromes
69. In general, factor analytic studies of the McGill Pain Questionnaire
- A) show one main factor, as postulated by Melzack
 - *B) support the three dimensional model of the author
 - C) suggest different factors for medical vs psychological pain
 - D) indicate that a person's perception of pain is the major factor
70. Among the criticisms of the McGill Pain Questionnaire:
- A) the scale is not sensitive to changes in pain over time
 - B) does not look at affective aspects of pain
 - C) little evidence of its validity
 - *D) the need for trained personnel to administer it
71. Type A behavior includes:
- *A) aggressiveness, competitiveness, and time pressure
 - B) social isolation and bizarre ideation
 - C) greater success as a therapist
 - D) a relaxed approach to life stresses
72. The original form of the Jenkins Activity Survey was based on:
- A) the MMPI
 - B) the DSM
 - *C) a structured interview
 - D) a life event inventory
73. On the Jenkins Activity Survey items are scored using
- *A) weights that reflect an item's ability to discriminate Type A from others
 - B) unitary weights of 0 and 1
 - C) arbitrarily assigned weights to reflect theoretical importance
 - D) weights on a standard Likert scale
74. The Jenkins Activity Survey was standardized on
- A) a large sample chosen on the basis of U.S. census data
 - *B) employed males in the age range of 44 to 64
 - C) both male and female college students
 - D) blue collar workers at risk for coronary heart disease

75. Test-retest reliability data on the Jenkins Activity Survey
- A) is rather poor due to the nature of coronary heart disease
 - B) is not appropriate since heart problems can appear suddenly
 - *C) reflects both change over time and differences between forms
 - D) is available only for atypical samples of college students
76. The Competency Screening Test which uses sentence completion stems
- *A) is not a projective technique because the respondent must provide appropriate responses
 - B) can be considered a projective device since it uses ambiguous sentence stems
 - C) is essentially a measure of psychological adjustment
 - D) can be used instead of the Rorschach to understand the client's psychodynamic functioning
77. The Referral Decision Scale is intended to be administered by
- A) psychologists
 - B) school teachers
 - *C) correctional officers
 - D) spouses who are abuse victims
78. In one study of the Referral Decision Scale a substantial number of false positives were identified. This probably means:
- A) prisoners who did not understand the legal proceedings
 - *B) relatively normal prison inmates identified as mentally ill
 - C) mentally ill prisoners who were not recognized as such
 - D) individuals who engaged in criminal activities but were not caught
79. The process of jury selection, in legal terms, is called:
- A) triage
 - B) competency to stand trial
 - C) criminal responsibility
 - *D) voir dire
80. In jury selection, one major finding is that jurors
- A) low on authoritarianism are more attentive to the evidence
 - *B) high on authoritarianism are more likely to convict
 - C) who understand voir dire are less judgmental
 - D) who prefer more conventional norms are less likely to convict
81. Part of the authoritarian personality includes
- A) a dislike for most authority
 - *B) hostility towards out-group members
 - C) a preference for ambiguity
 - D) the denial of hostile impulses
82. The Legal Attitudes Questionnaire yields scores on
- A) understanding of court procedures and knowledge of penalties
 - B) schizophrenia, bipolar disorder, and depression
 - *C) authoritarian, antiauthoritarian, and equalitarian
 - D) legal knowledge, and defendant's style of responding
83. Studies of the validity of the Legal Attitudes Questionnaire find
- A) little support for the Authoritarian scale
 - B) excellent support for all three scales
 - C) no support for all three scales
 - *D) little support for the Equalitarian scale

84. In the Narby et al. (1993) meta analysis of studies that looked at authoritarianism there was a
- A) substantial relationship with the defendant's mental health status
 - *B) very low relationship with defendant culpability
 - C) very high relationship with defendant culpability
 - D) significant relationship between authoritarianism and aggressiveness of the crime
85. The Myart v. Motorola (1966) case involved
- A) a psychologist's claim that the test was valid
 - *B) a qualifying test taken by a black applicant
 - C) the Wonderlic Personnel Test
 - D) an applicant to the University of Washington Law School
86. In the DeFunis v. Odegaard (1971) case which involved admission to a law school, the basic issue
- *A) involved a white applicant who charged racial discrimination
 - B) centered on a black applicant who was not promoted
 - C) focused on minimum competency legislation
 - D) involved truth-in-testing legislation
87. In the Target Stores (1989) case, the issue was that
- A) employers need to exercise reasonable care in hiring
 - B) race norming is prohibited
 - C) the Wonderlic Personnel Test was not valid
 - *D) MMPI and CPI items violated sexual, religious, and racial discrimination laws
88. A case that provided considerable impetus for the study of test bias:
- *A) Larry P. v. Riles (1972)
 - B) Diana v. State Board of Education of California (1970)
 - C) The Civil Rights Act of 1991
 - D) Debra P. v. Turlington (1981)
89. This case involved minimal competency legislation:
- A) DeFunis v. Odegaard (1971)
 - B) Albemarle Paper Company v. Moody (1975)
 - C) Larry P. v. Riles (1972)
 - *D) Debra P. v. Turlington (1981)
90. Which of these test items illustrates "ideas of reference"?
- A) people think I am strange
 - *B) others talk about me when I'm gone
 - C) I don't like social gatherings
 - D) others have it in for me

16 The Issue of Faking

1. As a response set, the tendency to gamble
 - A) decreases both reliability and validity
 - *B) increases reliability but decreases validity
 - C) increases both reliability and validity
 - D) decreases reliability but increases validity
2. The response set of inclusiveness would primarily affect tests like
 - *A) the Rorschach and the Adjective Check List
 - B) the MMPI and the CPI
 - C) the WAIS and the WISC
 - D) the Stanford-Binet and the McCarthy scales
3. Acquiescence tends to make
 - A) both true and false items more valid
 - *B) false items more valid and true items less valid
 - C) true items more valid and false items less valid
 - D) both true and false items less valid
4. Cronbach argued that response sets can be reduced by using
 - A) essay exams
 - B) items that are ambiguous
 - C) protective techniques
 - *D) multiple choice items
5. An aspect of the definition of malingering:
 - *A) intentional
 - B) unconscious
 - C) not related to gain
 - D) psychosomatic symptoms
6. In Lanning's (1989) study of CPI archives, the incidence of faking good ranged from
 - *A) 0 to 11%
 - B) 10% to 25%
 - C) 25% to 35%
 - D) none of these
7. One of the main conclusions from Lanning's (1989) study of CPI archives:
 - A) the highest rate of faking occurs among accused criminals
 - *B) the incidence of faking varies with different samples and different situations
 - C) faking is a highly frequent phenomenon
 - D) most psychiatric patients try to fake good

8. Osipov (1944) argued that malingerers
 - A) are like actors who know when and when not to fake
 - B) are usually quite subtle in their faking
 - C) cannot be identified because their faking is part of their illness
 - *D) take on a role and go to extremes
9. In the area of personality, what a person says or does is called _____, versus how a person acts which is _____.
 - A) true response; response set
 - B) style; content
 - C) response set; style
 - *D) content; style
10. Rorer (1965) distinguished between set which is _____, and style which is _____.
 - A) acquiescence; personality
 - *B) creating a certain image; selecting a response category
 - C) keyed true; keyed false
 - D) affective; cognitive
11. Faking good is seen as composed of two aspects:
 - A) a faking component and a personality aspect
 - B) a cognitive factor and an affective factor
 - *C) self-deceptive enhancement and impression management
 - D) response set and response style
12. From the viewpoint of self-deception vs. other deception, faking good involves
 - *A) more other deception
 - B) more self deception
 - C) equal amounts of both
 - D) neither one
13. Random responding
 - A) is considered a type of faking bad
 - *B) may reflect an honest mistake
 - C) is always conscious and intentional
 - D) is made up of impression management and self-deception
14. In the one group-one instruction method to develop a faking scale:
 - A) items are retained that show a high rate of endorsement
 - B) subjects are instructed to fake
 - *C) items are retained that show a low rate of endorsement
 - D) normal subjects and psychiatric patients are compared
15. In the one group-two instructions method to develop a faking scale:
 - A) the same items are administered twice, first with generic fake instructions and then with specific fake instructions
 - B) $\frac{1}{2}$ of the subjects are assigned to normal instructions and $\frac{1}{2}$ to "fake" instructions
 - C) the responses of "deviant" subjects are compared to those of normal subjects
 - *D) subjects answer an item pool under standard instructions and then retake it under faking instructions
16. The item "I experience visual hallucinations" is typically
 - *A) endorsed by malingerers, but not by psychiatric patients
 - B) endorsed by both malingerers and by psychiatric patients
 - C) not endorsed by malingerers or by psychiatric patients
 - D) endorsed by psychiatric patients but not by malingerers

The Issue of Faking

17. A suppressor variable is:
 - A) significantly associated with a criterion but not a predictor
 - B) significantly associated with both criterion and predictor
 - *C) significantly associated with a predictor but not a criterion
 - D) not significantly associated with either predictor or criterion
18. In the Hase and Goldberg (1967) study of the CPI scales that were developed by different methods:
 - *A) the stylistic scales were the least valid
 - B) the empirical method produced more valid scales
 - C) the least reliable scales were the rational ones
 - D) the original CPI scales were the most reliable
19. Dicken (1963) looked at the argument that stylistic scales on the CPI represented error variance; he concluded that
 - A) statistically correcting personality scores substantially increased validity
 - B) both reliability and validity were affected negatively
 - C) stylistic scales basically reflect acquiescence
 - *D) statistically correcting personality scores did not increase validity
20. Instructions that faking can be detected and might be punished
 - A) have no effect on the amount of faking
 - *B) do seem to reduce the amount of intentional distortion
 - C) is less efficient than using forced-choice items
 - D) has been thoroughly investigated as a way to reduce faking
21. Using forced-choice items that are matched for social desirability
 - A) is an excellent way to create faking resistant scales
 - B) represents one of the best techniques to discourage faking
 - C) assumes that the resulting scale is a suppressor variable
 - *D) results in scales that are still open to fakability
22. Filler items
 - A) are scored as a way of assessing faking
 - B) are typically used to measure the amount of faking
 - *C) are “extra” items in a scale, that are not scored
 - D) represent a very useful approach to control faking
23. In the Marin et al. (1992) study of responses given by Hispanics, Hispanics preferred _____ extreme responses and _____ acquiescent responses.
 - *A) more; more
 - B) less; more
 - C) more; less
 - D) less; less
24. In the Marin et al. (1992) study, less extreme and less acquiescent responses were related to _____ acculturation and _____ education.
 - A) lesser; lesser
 - B) lesser; greater
 - C) greater; lesser
 - *D) greater; greater
25. Symptom validity testing focuses on the notion that
 - A) faking is a symptom of maladjustment
 - *B) malingering results in scores below the chance level
 - C) malingering actually increases the validity of a test
 - D) impression management can be detected by the MMPI

26. In a multivariate test like the MMPI or CPI the order of items is randomized. The effects of such randomization
 - A) are generally positive, as shown by a large number of studies
 - B) have no impact on the validity of personality scales
 - *C) have not been investigated thoroughly
 - D) result in later-listed items that are less valid
27. Positional response bias refers to:
 - *A) selecting a response position (e.g. C) more often
 - B) labeling a scale with a nondescript title
 - C) preferring the left hand to respond to test items
 - D) a response set that reflects social desirability
28. The items on the CPI Sense of Well-being Scale were originally:
 - A) MMPI items endorsed with high frequency by neurotic patients
 - B) CPI items that were rarely endorsed by patients or controls
 - C) CPI items that showed large response shifts under instructions to fake
 - *D) MMPI items that discriminated between patients and dissimulators
29. In the Gough (1954) study of dissimulation, a cut-off score of 35 identified correctly about _____ % of the dissimulators.
 - A) 5
 - B) 25
 - C) 50
 - *D) 93
30. The “classic” study by Albert et al. (1980) with Rorschach protocols showed that clinicians
 - A) could not distinguish between Rorschach protocols given under normal instructions and instructions to fake
 - *B) could not distinguish Rorschach protocols of psychotic patients from those of normals instructed to fake
 - C) were fairly good at identifying Rorschach protocols that were faked
 - D) could not identify dissimulation on either the Rorschach or the Bender-Gestalt
31. In the Faust et al. study (1988) of faking on the WISC-R and the Halstead-Reitan:
 - *A) no clinicians attributed the results to malingering
 - B) malingering was easily detected by most experts
 - C) the authors concluded that faking did not occur
 - D) the Halstead-Reitan but not the WISC-R was sensitive to faking
32. Both Dicken (1963) and Rorer (1965) concluded that acquiescence
 - A) needed to be controlled in such inventories as the CPI
 - B) could easily be faked and not detected
 - *C) was not that important in personality inventories
 - D) was the most significant response set
33. On the MMPI the “validity” scales are in opposition to the _____ scales.
 - A) reliability
 - *B) clinical
 - C) internal consistency
 - D) faking
34. On the MMPI, the Cannot Say or ? scale
 - A) is a measure of suspiciousness
 - B) can vary only from 0 to 10
 - C) consists of 15 items that involve denial of aggression
 - *D) is the number of items omitted by the respondent

The Issue of Faking

35. One aspect of the Lie scale is that
- *A) all keyed responses are false
 - B) all of the items are endorsed by less than 10% of the norms
 - C) all keyed responses are true
 - D) all of the items concern aspects that most people would deny
36. Higher scores on the Lie scale are related to:
- A) better education and higher intelligence
 - *B) rigidity and conventionality
 - C) a high degree of morality and ethics
 - D) anxiety and depression
37. Keyed responses on the MMPI F scale:
- A) are correlated with the Schizophrenia scale
 - B) reflect anger and hostility
 - *C) reflect infrequency of response
 - D) are all false
38. A problem with the F scale is that elevated scores can also be associated with:
- *A) a "different" status like minority or maladjustment
 - B) all of the items are related to hostility
 - C) all the keyed responses are false
 - D) the scale is too short and hence unreliable
39. The K scale was developed by comparing _____ with those of normals.
- A) the responses of normal subjects instructed to fake
 - *B) responses of psychiatric patients whose MMPI profiles were normal
 - C) high scorers on the F scale
 - D) responses of normal subjects whose MMPI profiles showed maladjustment
40. Scores on the K scale are in general related to:
- *A) defensiveness
 - B) egotism
 - C) maladjustment
 - D) a moralistic view
41. The F-K index attempts to measure
- A) lack of self insight
 - B) positional response set
 - *C) fake bad
 - D) social desirability
42. We would expect that a normal subject who tries to fake bad on the MMPI:
- A) would underreport symptoms and difficulties
 - *B) would result in an elevated profile
 - C) would obtain an elevated schizophrenia scale only
 - D) would have an F-K score of -10 to -20
43. The Back-page Infrequency scale for the MMPI-2:
- A) consists of items dealing with negative behavior that most would admit
 - B) was developed along the lines of the K scale
 - *C) consists of items endorsed by a small percentage of the normative group
 - D) parallels the L scale of the original MMPI
44. The test-retest scale of the MMPI was based on the observation that:
- *A) the F scale is based on items from the first 300 items
 - B) the test-retest reliability of the MMPI scales is extremely high
 - C) the F-K index seems to be the best detector of faking
 - D) people who are consistent fake less

45. Using various scales to identify random response MMPI's, we would expect an accuracy rate of:
- A) less than 10%
 - B) about the level of chance
 - C) somewhere between 50% and 60%
 - *D) 90% or above
46. The MMPI subtle and obvious scales developed by Wiener (1948) were developed on the basis of:
- *A) the manifest content of the items
 - B) a statistical analysis of neurotic vs normal responses
 - C) their ability to discriminate between psychiatric and control samples
 - D) empirical correlations with a variety of scales
47. In the Austin (1992) study of fake good and fake bad instructions on the MMPI with college students:
- A) the Wiener subtle and obvious keys were the best indicators of fake good
 - B) the L and F scales were superior to the F-K index in identifying fake bad
 - C) none of the scales studied were able to detect faked profiles
 - *D) the F-K index showed a high rate of false positives for those instructed to answer honestly
48. How valid are MMPI validity scales as separate instruments?
- A) not very valid
 - *B) apparently quite valid
 - C) valid with college students but not psychiatric patients
 - D) valid, but with low test-retest reliability
49. What can we conclude about faking on the MMPI?
- A) all MMPI indices are equally successful
 - *B) normal individuals who fake psychosis can be readily detected
 - C) psychotics who exaggerate their condition cannot be detected
 - D) all of the above
50. The Sense of Well Being Scale of the CPI
- *A) is composed of MMPI items
 - B) originally was part of the Personality Research Form
 - C) is comparable in make-up to the test-retest scale of the MMPI
 - D) is a good example of a K corrected scale
51. A high scorer on the Sense of Well Being Scale could be described as:
- A) attempting to present a fake bad portrait
 - B) having a great degree of self-reliance and good interpersonal skills
 - C) socially awkward and conforming
 - *D) having good health and a positive outlook on life
52. The Good Impression Scale was developed by the _____ method.
- A) one group-one instruction
 - B) two group-two instructions
 - *C) one group-two instructions
 - D) two group-one instruction
53. Low scorers on the Good Impression Scale can be characterized as:
- *A) rebellious, critical, self-indulgent, and distrustful
 - B) considerate, cooperative, conscientious, and industrious
 - C) original, inventive, and creative
 - D) intelligent, independent, and good at problem solving

The Issue of Faking

54. In the Lanning (1989) study of regression equations to identify invalid CPI protocols, the results indicated:
- A) the use of the regression equations gave chance results
 - *B) the most difficult protocols to identify correctly were the fake good
 - C) the easiest protocols to identify as fake were computer generated
 - D) the three regression equations worked equally well
55. The Edwards (1957) Social Desirability Scale is composed of items from the
- A) CPI
 - B) Edwards Personal Preference Schedule
 - C) Personality Research Form
 - *D) MMPI
56. Zerbe and Paulhus (1987) argued that social desirability
- A) is a statistical contaminant that can be held in check
 - B) is of little importance in personality testing
 - *C) may indeed be related to the construct being tested
 - D) of an item is correlated to rate of endorsement
57. There are two ways of interpreting social desirability:
- A) as a way of faking bad and faking good
 - B) as an individual dimension or as a group consensus
 - C) as affecting reliability or affecting validity
 - *D) as a contaminant and as a personality dimension
58. McCrae and Costa (1983) presented evidence to support the notion that:
- A) individuals who score higher on social desirability are faking good
 - B) scales of social desirability measure individual differences in SD
 - *C) correcting for social desirability does not increase validity
 - D) spouse ratings as a criterion are not very valid
59. The Marlowe-Crowne SD Scale consists of items that
- A) reflect psychopathological content
 - *B) describe culturally approved behaviors with a low probability of occurrence
 - C) show very low endorsement for social desirability
 - D) that are presented in forced-choice pairs and equated for SD
60. In the Holden and Fekken (1989) study of Canadian college students, the two SD scales that correlated most highly with each other:
- *A) Edwards vs. Jackson
 - B) Edwards vs. Marlowe-Crowne
 - C) Jackson vs. Marlowe-Crowne
 - D) Marlowe-Crowne vs. Gough
61. In regards to reducing social desirability, using a forced-choice format:
- *A) does not eliminate social desirability
 - B) removes the self-deception component but not impression management
 - C) works very well if the subject tries to consciously fake
 - D) is probably the best available approach
62. To reduce social desirability we can use neutral items. This suggestion is:
- A) an excellent one that is used in a number of tests
 - B) well supported by the literature
 - *C) not a good one since such items are rare
 - D) applicable only to cognitive tests

63. The “bogus pipeline” technique involves a situation where
- A) the subject is asked to fake in a positive direction
 - *B) the subject believes that truthful responding can be detected
 - C) neutral items are presented and the subject answers in a socially desirable manner
 - D) the items used are quite socially obvious
64. In retaining items for their scale, Crowne and Marlowe
- A) used the F-K index as their main criterion
 - *B) kept items that discriminated between low and high scorers
 - C) used only MMPI items with “subtle” content
 - D) retained items where there was disagreement as to their SD
65. Scores on the Edwards SD Scale correlate negatively with most MMPI clinical scales. This has been interpreted as indicative that the Edwards Scale:
- A) is not a good measure of social desirability
 - *B) measures the willingness to endorse neurotic symptoms
 - C) is essentially a measure of interpersonal skills
 - D) is a good predictor of early termination from psychotherapy
66. From a psychometric point of view, a “defect” in the Marlowe-Crowne scale is that:
- A) both attribution and denial items are keyed false
 - B) it does not have high correlations with the Edwards
 - *C) attribution items are keyed true, and denial items are keyed false
 - D) all of its items are loaded on a factor of psychopathology
67. The net effect of acquiescence is that
- A) the reliability and validity of true keyed items will be greater than false keyed items
 - B) reliability will increase but validity will decrease
 - C) test-retest reliability will decrease
 - *D) the reliability and validity of false keyed items will be greater than true keyed items
68. A problem with most scales that measure acquiescence:
- *A) substantial correlations with social desirability
 - B) low internal consistency
 - C) poor test-retest reliability
 - D) significant correlations with MMPI clinical scales
69. Jackson (1967) argued that acquiescence
- *A) should be controlled when the test is being developed
 - B) is of little concern and may not even exist
 - C) can be easily controlled through factor analysis
 - D) can be eliminated by keying all the responses false
70. In writing test items most authors attempt to balance positively versus negatively worded items. There is evidence that
- *A) negative items may lower reliability and possibly validity
 - B) such an approach is the best way to control for acquiescence
 - C) this practice increases reliability but not validity
 - D) such an approach does not control acquiescence
71. At least for the MMPI we can conclude that
- A) indices like the F-K control acquiescence well
 - *B) acquiescence does not seem to play a major role
 - C) acquiescence can be eliminated through appropriate instructions
 - D) the clinical scales are significantly affected by acquiescence

The Issue of Faking

72. Scores on the Edwards Social Desirability Scale correlate significantly with scores on most MMPI scales. This may well be because:
 - A) most MMPI scales are heavily influenced by social desirability
 - B) both the MMPI and the Edwards measure psychopathology
 - C) both the MMPI and the Edwards measure acquiescence
 - *D) the items on the Edwards are MMPI items
73. In the study by Schretlen and Arkowitz (1990) of prison inmates instructed to fake, the results showed that:
 - A) the F-K index on the MMPI was the best predictor of faking
 - *B) discriminant analyses could correctly classify 92 to 95% of subjects
 - C) the CPI validity scales were the most effective measures
 - D) psychiatric patients exhibited more acquiescence than prisoners
74. An example of an unspecified frequency statement:
 - A) all depressed individuals attempt suicide
 - *B) depressed individuals are suicidal
 - C) depression leads to suicide
 - D) most depressed individuals may be suicidal
75. In the Furnham (1990) study of the fakeability of the Myers-Briggs:
 - A) neither fake good nor fake bad instructions altered test profiles
 - B) fake good instructions increased mean scores on few scales
 - C) the results were the same as with the 16 PF
 - *D) both fake good and fake bad instructions produced higher mean scores
76. The initial validation studies of the BDI were conducted on:
 - A) patients, therefore highly motivated to fake
 - *B) patients, therefore not highly motivated to fake
 - C) college students, therefore not highly motivated to fake
 - D) community adults with little incentive to fake
77. In the study by Kornblith et al. (1984) comparing BDI results under standard instructions versus various conditions, the results suggest that:
 - *A) item sequence does not affect obtained scores
 - B) different item sequences result in drastic changes in means
 - C) instructions that emphasize depression resulted in substantial mean changes
 - D) students who were more depressed perceived the BDI as more threatening
78. In the Dahlstrom et al. (1990) study of the BDI and two alternate forms of reversed and random sequence, the results indicated that:
 - A) there were no significant mean differences among the 3 forms
 - B) the reverse sequence substantially increases criterion validity
 - *C) the random order form resulted in a higher mean score
 - D) the standard version was the most valid
79. The Symptom Overendorsement Scale of the Psychological Screening Inventory:
 - *A) assesses the extent to which a normal individual endorses psychopathology
 - B) was developed by comparing college students and psychiatric patients
 - C) can be considered comparable to the K scale of the MMPI
 - D) assesses the extent to which a schizophrenic endorses psychopathology
80. The Endorsement of Excessive Virtue Scale of the Psychological Screening Inventory is made up of items:
 - A) judged to represent superior adjustment, and endorsed by healthy subjects
 - B) that discriminate between psychiatric patients and well adjusted persons
 - *C) judged to represent excessive virtue, and endorsed by few individuals
 - D) all of the above

81. Evidence for the validity of the four deception scales on the PSI:
 - A) the four scales correlate substantially and positively with each other
 - B) the scales do not correlate with the MMPI validity scales
 - C) high scorers on the four scales tend to be maladjusted
 - *D) simulated deception significantly altered mean scores
82. In Harrington's (1975) study of the ACL creativity scale:
 - *A) results under the creative instructions were more valid
 - B) the results indicated a strong impression management component
 - C) scores on the scale correlated significantly with social desirability
 - D) the F-K index was able to detect most faking
83. In the Ryan and Sackett (1987) study of the fakability of an integrity test, the results showed that:
 - A) integrity scales can be easily faked
 - *B) subjects responding as if applying for a job were basically honest
 - C) the "admission" component was the most susceptible to faking
 - D) subjects instructed to respond as a job applicant were the least honest
84. As to intelligence and integrity test scores:
 - *A) there seems to be no relationship
 - B) studies have found a moderate positive relationship
 - C) more intelligent applicants lie less frequently
 - D) there appears to be a curvilinear relation between the two
85. In the Cunningham et al. (1994) study of impression management on the Reid Report, the authors concluded that integrity tests
 - A) have rather poor validity
 - B) are highly susceptible to faking
 - C) basically measure social desirability
 - *D) possess predictive validity
86. In the option-keying strategy with biodata items the scoring weights:
 - A) go from 1 to 5 as in a typical Likert response option
 - *B) reflect which alternatives differentiate the contrasted groups
 - C) are applied in an identical manner to the item-keying strategy
 - D) reflect the judgment of a group of experts
87. Biodata items scored with the option-keying procedure:
 - A) are less valid
 - *B) are less amenable to faking
 - C) are more valid
 - D) are more amenable to faking
88. In the Cohen and Lefkowitz (1974) study of a biodata questionnaire and the MMPI K scale, the results supported the notion that:
 - A) high K scorers faked the biodata questionnaire
 - B) scores on the K scale are related to psychopathology
 - C) the F-K index was more valid than the K by itself
 - *D) high K scorers were basically well adjusted with middle class values
89. The Kahn et al. (1988) study of faking on Rorschach protocols indicated that:
 - *A) a computer scoring system was not very accurate
 - B) when psychotic patients are asked to respond "normally" they do so
 - C) when subjects are instructed to malingering they give more responses
 - D) if college students are asked to fake, they lack the motivation to do so

The Issue of Faking

90. In the study by Timmons et al. (1993) of a sentence completion test, they found that:
 - A) exaggerated confidence in one's doctor was the main factor
 - B) the MMPI L scale was the best indicator of faking
 - *C) the 3 factor scales they developed had both convergent and discriminant validity
 - D) when subjects were instructed to malingering they gave shorter responses
91. In assessing malingering on neuropsychological tests, the usual research procedure is to:
 - A) instruct patients with brain injury to malingering
 - *B) instruct normal subjects to malingering
 - C) compare the performance of high and low scorers on the L MMPI scale
 - D) follow up testing of a patient with an interview to detect faking
92. In studies of faking using the Wechsler Memory Scale or similar instruments, the results show that:
 - *A) discriminant equations are quite accurate
 - B) such instruments are easily fakable
 - C) there are too many false positives to obtain valid results
 - D) the validity scales from the CPI can accurately identify malingering
93. A number of studies have shown that malingered and valid WAIS profiles
 - A) cannot be distinguished with any degree of accuracy
 - B) differ from each other on Performance subtests only
 - C) can be discriminated by the F-K index
 - *D) can be distinguished on the basis of subtest patterns
94. In regards to malingering on neuropsychological tests, we can conclude that:
 - A) normal subjects, if instructed, can simulate brain damage
 - B) clinicians have poor rates of detection of malingering
 - C) discriminant functions and other procedures work relatively well
 - *D) all of the above
95. On the Halstead-Reitan Battery we can probably conclude that malingering
 - A) is generally not detectable
 - *B) is detectable by using complex statistical procedures
 - C) can be detected only when normal subjects are tested
 - D) is a major problem that has not yet been resolved
96. The first widely used test anxiety instrument was the
 - A) Test Anxiety Questionnaire
 - B) Test Anxiety Scale
 - C) Test Anxiety Inventory
 - *D) Test Anxiety Scale for Children
97. In the state-trait model of test anxiety
 - A) trait anxiety is seen as a reaction to a particular situation
 - B) state anxiety progresses to trait anxiety
 - *C) state anxiety is seen as transitory
 - D) state anxiety is seen as a relatively stable characteristic
98. In regards to test anxiety and defensiveness in children:
 - A) defensiveness and the lie scale do not correlate
 - *B) highly defensive children admit to less anxiety
 - C) there is no relationship between defensiveness and anxiety
 - D) highly defensive children admit to more anxiety
99. In general, test anxiety seems to be composed of two separate aspects:
 - A) trait anxiety and phobia
 - *B) worry and emotionality
 - C) school phobia and fear of failure
 - D) self anxiety and other anxiety

100. Testwiseness refers to:
- *A) individual differences in test-taking skills
 - B) lack of test anxiety
 - C) intellectual competence on tests
 - D) the ability to guess correctly
101. Testwiseness can be analyzed into two components:
- A) state versus trait
 - B) fake good versus fake bad
 - *C) independent of the test and dependent
 - D) cognitive and affective
102. Testwiseness can be substantially reduced by:
- A) instructions not to guess
 - *B) well written items
 - C) a correction formula
 - D) using the L MMPI scale

17 The Role of Computers

1. One of the first uses of computers in psychological testing was to:
 - *A) score tests that previously had been hand or machine scored
 - B) analyze MMPI protocols
 - C) administer tests to a client
 - D) print out a test score interpretation
2. The initial computer programs to score the MMPI yielded
 - A) a set of raw scores only
 - *B) a set of standard scores and some simple descriptive statements
 - C) an interpretive summary based on configural patterns
 - D) a rather sophisticated interpretive summary
3. Commercially available computer interpretive programs for personality tests typically
 - A) are supported by a wide variety of consumer information
 - B) describe how the programs work in such references as the MMY
 - C) present raw or standard scores only with no interpretation
 - *D) do not give information on how the program works and its validity
4. Configural scoring means that
 - A) the entire test profile, as in an MMPI, is considered impressionistically
 - B) the emphasis is placed on graphing the results
 - *C) the answer to a set of items is considered as a unit
 - D) items are scored as usual, but only items keyed false are considered
5. One major advantage of the computer in test administration:
 - A) can handle any length test
 - B) can be programmed to present items at a regular pace
 - *C) demands an active response
 - D) can display practice items
6. From a theoretical and classical perspective, equivalence means
 - *A) identical means, variances, and patterns of correlations
 - B) that sources of generalizability can be identified
 - C) that the two test forms are measuring the same variable
 - D) equivalence in construct validity
7. In a mastery test like an algebra achievement test
 - A) the focus is usually on knowing the weaknesses and strengths of the subject
 - *B) items are usually presented sequentially
 - C) constructed-response items are typically used
 - D) items are usually presented in a branching format

8. An example of a constructed-response item:
 - A) multiple choice
 - B) true-false
 - C) matching
 - *D) essay
9. Most studies of the comparability of p-p tests with their cf versions indicate
 - A) relatively little equivalence
 - *B) a great deal of equivalence
 - C) lack of equivalence on such tests like the MMPI
 - D) equivalence within a sample but not across samples
10. Finegan and Allen (1994) compared a variety of p-p questionnaires with their cf counterparts in Canadian college students. Their results support the conclusion that:
 - A) equivalence is difficult to demonstrate
 - B) equal means do not necessarily reflect equivalence
 - C) attitudinal variables should be subsumed under personality
 - *D) there may be an interaction of format with personality variables
11. In the study of the Raven's Progressive Matrices (Watts et al. 1982) with both p-p and cf versions:
 - A) equivalence was not obtained, but separate norms were not needed
 - *B) equivalence was not obtained and separate norms were needed
 - C) equivalence was obtained so separate norms were not needed
 - D) the authors concluded that the cf version was not sufficiently reliable
12. In regards to equivalence of reliability
 - A) there are few studies and they suggest this may be a difficulty
 - B) we can assume equivalence for most tests
 - *C) there are few studies but they support such equivalence
 - D) the computer format seems to increase reliability
13. In regards to equivalence of speeded tests
 - A) most such tests seem to be highly equivalent
 - *B) we would expect higher scores on the cf version
 - C) the reliability is higher on the cf version
 - D) we would expect lower scores on the cf version
14. One possible advantage of cf of speeded tests:
 - A) reliability seems to increase given the standardized presentation
 - B) subjects prefer the ease of answering with computer keys
 - C) mean scores are typically higher reflecting a better performance
 - *D) both time elapsed and items attempted can be tracked
15. Do examinees like the computer? Most studies show that
 - A) such a question cannot be answered experimentally
 - B) examinees respond negatively
 - *C) the answer seems to be yes
 - D) there is a gender difference, with males responding positively
16. In the Knight et al. (1988) study of the computer as a test preparation mentor, the results showed that:
 - A) the computer group scored significantly higher on all ACT subjects
 - *B) the computer can help in some areas but not in other areas
 - C) the computer group did equally well as the control group
 - D) high school students lacked the required maturity to benefit

The Role of Computers

17. In the Lucas et al. (1977) study of alcoholic patients:
 - A) lesser amounts of alcohol consumption were reported via computer
 - B) computer administered interviews yielded essentially identical results to standard interviews
 - C) psychiatric interviews were seen as more aversive by the patients
 - *D) greater amounts of alcohol consumption were reported via computer
18. In making course examinations available on computers to students, a major issue is that of
 - *A) test security
 - B) reliability
 - C) equivalence
 - D) validity
19. CBTI stands for:
 - A) clinical based test interpretation
 - *B) computer based test interpretation
 - C) computer backed test inventory
 - D) correct before test instructions
20. In terms of amount of information, the least amount is presented by a:
 - A) CBTI report
 - B) screening report
 - *C) descriptive report
 - D) consultative report
21. The original Mayo Clinic MMPI program yielded a:
 - *A) descriptive report
 - B) consultative report
 - C) diagnostic report
 - D) screening report
22. If we sent an MMPI for scoring by a professional service, we would expect to receive a _____ report.
 - A) descriptive
 - *B) consultative
 - C) screening
 - D) diagnostic
23. A major problem with the actuarial cookbooks for the MMPI:
 - A) they were too complicated to use by the typical clinician
 - B) the classificatory systems used had low reliability
 - C) their validity was difficult to document
 - *D) a large number of MMPI profiles could not be classified
24. Meehl's call for a good cookbook reflected the assumption that
 - A) clinical data was more valid than actuarial data
 - *B) MMPI profiles could be classified into distinct categories
 - C) the clinical scales of the MMPI could be changed to regression equations
 - D) MMPI profiles are unique since clients are unique
25. Clinicians attempt to take into account the uniqueness of their client. The result is:
 - A) increased validity of their clinical reports
 - B) that such uniqueness can be captured by MMPI profiles
 - *C) decreased validity of their clinical reports
 - D) that their validity increases but their reliability decreases

26. In the Kleinmuntz (1969) study of experienced MMPI clinicians, the results were that:
- *A) a computer program did as well as the best MMPI clinician
 - B) a computer program did better than the best clinician
 - C) none of the clinicians did as well as the computer program
 - D) most clinicians did better than the computer program
27. In regards to the reliability of CBTI's:
- A) the same MMPI scored by different services will yield identical results
 - *B) the same MMPI scored repeatedly by the same service will yield identical results
 - C) different MMPI's scored by the same service will yield identical results
 - D) test-retest reliability is relatively high but others are low
28. Consumer satisfaction studies of CBTI typically involve
- A) asking clients how well the report matches their personality
 - B) asking clients how satisfied they were with the computer testing
 - C) having experts match CBTI's with clinical reports
 - *D) asking clinicians to rate the accuracy of such reports
29. Studies that compare judgments based on CBTI's versus judgments based on direct knowledge of the client yield a median correlation of about:
- A) .10
 - *B) .30
 - C) .50
 - D) .80
30. Both Dr. Jones and Dr. Smith do a Q sort to describe the same client. Dr. Jones bases his knowledge solely on a CBTI while Dr. Smith is the client's therapist. The two Q sorts correlate .92. We would conclude that:
- *A) the results might reflect incorrect judgments by both clinicians
 - B) such results show the value of Q sorts as a validity method
 - C) CBTI's can be quite valid
 - D) Dr. Smith has a better diagnostic picture than Dr. Jones
31. Seeing that a report is computer generated may make the results more acceptable as correct and objective. Such a statement is:
- A) partially supported in a study of school psychologists
 - *B) not supported by the available literature
 - C) correct when the perceiver is a lay person
 - D) well supported by the available literature
32. Two potential limitations of CBTIs have been identified. The first has to do with:
- A) validity
 - B) reliability
 - C) generality
 - *D) misuse
33. The Aunt Fanny report refers to:
- *A) a report that is so general it could apply to anyone
 - B) a report that is based on stereotypes of older persons
 - C) a report that is quite accurate in describing a unique person
 - D) a screening report used to identify clients with specific problems
34. One of the major problems in studying the _____ of CBTIs is the _____.
- A) validity; Aunt Fanny report
 - B) reliability; Aunt Fanny report
 - *C) validity; Barnum effect
 - D) reliability; Barnum effect

35. A "Barnum" version of a CBTI would be made up of statements that
 - A) focus on psychopathology
 - B) contain a lot of professional jargon
 - C) are not applicable to the particular client
 - *D) are generic and true of most people
36. In the O'Dell (1972) study of three forms of CBTIs, the most accurate version was judged to be the
 - A) computer generated one
 - *B) Barnum one
 - C) prosecuting attorney one
 - D) "real" one
37. In the O'Dell (1972) study of three forms of CBTIs, the most disliked version was judged to be the
 - A) computer generated one
 - B) Barnum one
 - *C) prosecuting attorney one
 - D) "real" one
38. The major point of the O'Dell (1972) study of various forms of the CBTI:
 - *A) accuracy must be judged against a standard
 - B) various forms are equally valid
 - C) the Barnum effect can be a useful one
 - D) reliability reflects the length of the report
39. Bogus reports can be made up in many ways such as:
 - A) having inexperienced students write such reports
 - *B) using an average test profile
 - C) having the subject indicate which statements are correct
 - D) using the profile of a famous person
40. The few studies that compare a CBTI to the traditional clinical report can be criticized because:
 - A) only one clinician is involved
 - B) the CBTI used is an Aunt Fanny report
 - *C) they use one case only
 - D) accuracy judgments are made by inexperienced college students
41. Computer scoring of neuropsychological tests would not necessarily provide a major advantage because:
 - A) such tests are usually not scored but are impressionistically evaluated
 - *B) such scoring is usually quite simple
 - C) such scoring is typically complex and time consuming
 - D) computer scoring programs are difficult to design
42. Test administration of neuropsychological tests
 - *A) may be difficult because of the nature of such assessment
 - B) is very useful and has been implemented for some time
 - C) has the advantage that clients can follow directions
 - D) is at about the same stage of development as the MMPI
43. The Heaton et al. (1981) study compared the accuracy of a computer program versus skilled psychologists in assessing brain damage. The results showed that:
 - A) the computer program was more accurate than the clinicians
 - B) the computer program was better at determining the issue of laterality
 - C) clinicians were better in one of the three major questions
 - *D) clinicians were better in predicting the three major questions

44. In neuropsychological assessment there are three major issues. Which one does NOT belong?
- *A) reliability of brain damage
 - B) presence of brain damage
 - C) localization of brain damage
 - D) chronicity of brain damage
45. The early investigations of computer programs for the Halstead-Reitan Neuropsychological test battery concluded that such programs
- A) were simply too unreliable and invalid
 - *B) could not be as accurate as experienced clinicians
 - C) could assess the presence but not the localization of brain damage
 - D) had great promise and showed both sensitivity and utility
46. Adams et al. (1984) in their study of computer programs to assess brain-damaged patients concluded that:
- A) patients with left hemisphere damage could be correctly identified
 - B) one of the three computer programs was potentially quite useful
 - *C) all three computer programs were inadequate
 - D) the WAIS index was superior to the Halstead-Reitan
47. One solution to the _____ dilemma involves _____. .
- *A) bandwidth-fidelity; adaptive testing
 - B) criterion vs. construct validity; CBTIs
 - C) Aunt Fanny; adaptive testing
 - D) running average; a large item bank
48. An early example of adaptive testing can be found on the
- A) MMPI
 - B) ACL
 - C) Raven's PM
 - *D) Stanford-Binet
49. True adaptive testing could not be used with the MMPI or CPI because such inventories
- A) are not based on item response theory
 - *B) have multidimensional scales
 - C) use true-false items
 - D) are too lengthy
50. Branching involves
- *A) using a small subset of items to determine whether to continue testing or move to another variable
 - B) is identical to the countdown method but applied to cognitive tests
 - C) a number of separate inventories that are merged into one test
 - D) adaptive testing of adults rather than children
51. In the Roper et al. (1995) study of the countdown method with the MMPI, the results showed that
- A) clinicians were much more comfortable with the standard MMPI
 - *B) a high degree of comparability among the three MMPI forms
 - C) the countdown method was not particularly useful
 - D) CBTIs based on the countdown method were most valid
52. A "library" is to a "bank" as
- A) personality inventories are to cognitive tests
 - *B) CBTIs are to tests
 - C) adaptive testing is to computerized testing
 - D) MMPI is to CPI

53. The future use of computers in the testing process
- A) will probably mean higher unemployment rates for psychologists
 - B) will mean that professionals other than psychologists will develop tests
 - *C) might mean the use of “assistant psychometricians”
 - D) might result in instruments with lower reliability and validity
54. Computerized testing with severely disabled clients
- A) is simply not feasible with our present technology
 - *B) might be more possible than p-p testing
 - C) is too time consuming to be feasible
 - D) yields rather unstable results
55. In the study by Wise and Wise (1987) of immediate feedback:
- A) the three versions of the test did not show equivalency
 - B) immediate feedback resulted in higher state anxiety for all subjects
 - C) delayed feedback was preferred by most teachers
 - *D) there seemed to be an interaction between level of achievement and immediate feedback
56. In the CBTI Guidelines a test user is defined as:
- *A) a qualified professional with knowledge of tests and related topics
 - B) any professional, like physicians and lawyers
 - C) psychologists only, who have the appropriate training
 - D) anyone who because of their position has clients
57. The CBTI Guidelines were developed by:
- A) test publishers like Consulting Psychologists Press
 - *B) professional organizations like the APA
 - C) individual professionals who use tests
 - D) test authors who produce tests
58. The CBTI Guidelines that cover interpretation are
- A) centered primarily on standardization of procedure
 - *B) consistent with the view that CBTIs are professional-to-professional communications
 - C) are based on the notion that professional judgment has low validity
 - D) directed at the various methods to obtain equivalency
59. Ziskin and Faust (1988) reviewed some 1400 studies on the reliability and validity of clinical evaluations conducted for legal purposes. They concluded that:
- A) while reliability is quite solid, validity is questionable
 - B) psychological theory is quite advanced but measurement is not
 - *C) the studies cast doubt on the reliability and validity of such evaluations
 - D) the problem lies with the tests rather than the test users
60. The “Air Defense Game” involving a simulated radar screen was used as an example of:
- *A) a potentially useful test developed specifically for computer administration
 - B) a p-p test that was successfully translated into computer format
 - C) how validity increases with computer administration
 - D) the gender difference in a variety of areas
61. Barrett et al. (1982) reported on a battery of tests, for example one using equilateral triangles. The results were said to be:
- A) exciting because of high reliability and moderate validity
 - B) a statistical artifact because of the nature of the test
 - C) a good example of the application of CBTI Guidelines
 - *D) disappointing because of low test-retest reliability

62. The study by Veldman (1967) of a computerized sentence completion test
- *A) has at present little utility
 - B) shows how such a test could be used by clinicians
 - C) points to the fact that projective techniques have low validity
 - D) shows that the reliability of the computerized version is lower
63. In the study by Steer et al. (1994) of the Beck Depression Inventory with psychiatric inpatients:
- A) BDI scores were significantly related to demographic variables
 - B) anxiety patients scored higher than depressed patients
 - C) the coefficient alpha was below .70
 - *D) the reliability and validity of the cf were comparable to the p-p
64. In terms of behavioral assessment, microcomputers
- A) have not found wide application in such assessment
 - *B) have basically revolutionized such assessment
 - C) are simply of no use to most behaviorists
 - D) have been applied in a very limited way
65. In the Dunnet et al. (1972) study of a cf MMPI and response latencies it was found that
- A) males tended to answer more rapidly than females
 - B) response latency was related to higher maladjustment
 - *C) response time was a function of item length
 - D) response latency was significantly related to item content
66. In the Cummings (1986) series of studies on CBTIs given (or not given) to physicians, the results indicated that:
- A) simpler reports were preferred by most physicians
 - *B) as the complexity of the report increased, there were more referrals
 - C) the referral rate to psychotherapy remained unchanged
 - D) the CBTIs were more valid than the physicians' judgment
67. The study by Johnson and White (1980) on the computerized version of the Wonderlic, suggests that:
- A) degree of familiarity with computers is not correlated with performance
 - B) college students do better than older persons because they are computer literate
 - C) the two versions lacked equivalence
 - *D) familiarity with computers can increase a person's score
68. The Dutch test for pharmaceutical chemist's assistants was used to illustrate:
- *A) what a future interactive computer-video test might look like
 - B) how voice output can be used to present instructions
 - C) that constructed response items are more reliable than selected response items
 - D) how computer scoring is much more reliable

18 Testing Behavior and Environments

1. In classical assessment it is hypothesized that:
 - *A) test performance is a sign of an underlying dimension
 - B) only personality variables can be measured
 - C) test responses and the underlying trait are identical
 - D) behavior is a function of the stimuli present
2. Behavioral assessment covers
 - A) only physiological-emotional behavior
 - *B) all activities, even cognitive and verbal
 - C) what can be observed by another person
 - D) such aspects as traits and motives
3. Direct observation is preferred by behaviorists because
 - A) they are easy to implement in a research setting
 - B) they distrust self-report
 - C) such observations are most reliable
 - *D) they are empirically verifiable and not inferential
4. A major problem with observation of behavior:
 - A) what theoretical rationale can we give?
 - B) are obtrusive observations valid?
 - *C) is interrater reliability sufficiently high?
 - D) how do we determine alternate forms reliability?
5. A major problem with self-monitoring:
 - *A) the accuracy of the procedure
 - B) cooperation from the subject
 - C) unobtrusive observations have low reliability
 - D) the subject has to be instructed
6. In the context of behavioral assessment, reactivity means that
 - *A) the behavior may change because it is being monitored
 - B) a particular behavior may be quite unstable
 - C) direct observation can be disruptive
 - D) there is little or no rapport between subject and observer
7. A basic assumption of role playing:
 - A) we are defined by the roles we play
 - B) not a test, so reliability is of no concern
 - C) people behave in stable ways across different roles
 - *D) how the client responds reflects real life behavior

8. The most frequently used technique in behavioral assessment:
 - A) self monitoring
 - B) checklists and rating scales
 - *C) interviews
 - D) role playing
9. In the Sarbin (1943) study of academic success
 - A) interviews were found to be highly unreliable
 - B) academic success was higher for women than for men
 - *C) interview information added nothing to the prediction
 - D) interviews were used as a screening procedure
10. In behavioral assessment, reliability of interviews is often assessed by:
 - A) test-retest
 - B) split-half
 - *C) interrater agreement
 - D) internal consistency
11. In structured interviews _____ validity seems particularly relevant.
 - *A) content
 - B) construct
 - C) criterion
 - D) primary
12. From a psychometric point of view
 - A) interviews are the best of the indirect procedures
 - *B) interviews present a number of potential problems
 - C) interviews are generally considered reliable and valid
 - D) it is not important to determine the reliability of interviews
13. Behavioral self-report scales differ from other self-report scales in that
 - *A) they focus on behavior
 - B) they are typically shorter
 - C) their reliability is well established
 - D) they represent an unobtrusive technique
14. Behaviorists are somewhat suspicious of checklists and rating scales because:
 - *A) they are indirect measures of behavior
 - B) most come from a theoretical framework
 - C) they are heavily influenced by social desirability
 - D) they are newer measures that have not been validated
15. From a behavioral assessment point of view, self-report
 - A) should be the primary way to assess
 - *B) can have several advantages
 - C) has the highest inter-rater reliability
 - D) is the easiest to evaluate objectively
16. A major difference between traditional and behavioral assessment:
 - A) intervening variables are central to behavioral assessment
 - B) both use intervening variables, but traditional assessment sees them as direct methods
 - *C) behavioral assessment does not use intervening variables
 - D) traditional assessment is suspicious of intervening variables
17. In regards to specificity and generality
 - *A) traditional testing focuses on generality
 - B) behavioral assessment is suspicious of both
 - C) traditional testing focuses on specificity
 - D) their validity is difficult to support

18. Content validity is very important in behavioral assessment because
 - A) the other types of validity are not appropriate
 - B) adequate sampling is of little concern to behaviorists
 - *C) the observations made must be representative of the behavior
 - D) the focus is on the individual rather than the group
19. The effectiveness of a behavior change as assessed by its utility:
 - A) consumer satisfaction
 - *B) social validity
 - C) adequate sampling
 - D) concurrent validity
20. Scorer generalizability refers to:
 - A) stability over time
 - *B) the degree of correspondence between two observers
 - C) internal consistency among different observers
 - D) what ordinarily would be called construct validity
21. An example of time generalizability:
 - A) how well a person uses their time in different settings
 - B) how long it takes to collect specific observations
 - C) a K-R or alpha coefficient
 - *D) test-retest reliability
22. From a psychoanalytic perspective, self-reports are considered
 - A) both reliable and valid
 - *B) reliable but not valid
 - C) neither reliable nor valid
 - D) not reliable but valid
23. A typical behavioral assessment scale would
 - A) have high inter-rater reliability
 - B) have very little reactivity
 - *C) not yield a total score
 - D) focus on internal consistency
24. The sensitivity of a behavioral assessment technique refers to:
 - *A) how well the technique reflects changes in behavior
 - B) the degree to which chance agreement is accounted for
 - C) how measurement can change the actual behavior
 - D) the length of time needed to collect observational data
25. If there is reactivity present, the result will be:
 - A) a greater incidence of behavior when observed
 - B) a smaller incidence of behavior when observed
 - *C) a change in behavior when observed
 - D) no change in behavior when observed
26. The Kappa statistic is used to assess:
 - A) chance agreement
 - *B) interobserver agreement
 - C) internal consistency
 - D) observed agreements
27. Behavioral checklists and rating scales
 - A) are essentially identical
 - *B) differ as to how global the judgment is
 - C) measure personality and behavior respectively
 - D) are less valid than interviews

28. If a behavior to be rated is very specific then
A) inter-observer reliability may be quite low
B) the use of a control group is required
C) the Kappa coefficient would be useful
*D) test-retest reliability may be inappropriate
29. From the viewpoint of behavioral assessment different ratings provided by teachers and by mothers:
A) are a good way to measure interrater reliability
*B) may be legitimate sources of information
C) should be almost identical
D) suggest that the rating scale may not be valid
30. Response bias means that
A) the observer is not being fully objective
*B) different item arrangements may elicit different responses
C) the results are confounded by chance agreement
D) different item arrangements may elicit the same response
31. From a behavioral assessment point of view, specificity
A) should be avoided where possible
B) results in lowered reliability
C) can be controlled by the Kappa coefficient
*D) is a desirable quality of items
32. The Children's Inventory of Anger uses _____ as items.
*A) mini-vignettes
B) adjectives
C) true-false statements
D) happy faces
33. The reliability of the Children's Inventory of Anger
A) is low as expected of this type of instrument
B) is about average for a personality test
*C) appears to be quite substantial
D) has not been investigated
34. Validity evidence for the Children's Inventory of Anger:
A) substantial correlations with MMPI scales
B) parental reports agreed with teacher reports
C) emotionally disturbed children scored higher than controls
*D) scale reflects changes due to therapeutic treatment
35. The Daily Child Behavior Checklist presumably uses items
A) that describe interpersonal conflict situations
B) all of which focus on negative behaviors
*C) that are objective and focus on factual events
D) selected on the basis of factor analysis
36. The authors of the Daily Child Behavior Checklist concluded that their instrument had satisfactory reliability
A) and the evidence supports their conclusion
B) even though no data is presented
C) but only of the internal consistency kind
*D) but the test-retest coefficients are below .70

37. In general the validity of the Daily Child Behavior Checklist
- *A) seems rather marginal
 - B) is excellent and covers a variety of bases
 - C) is based primarily on significant correlations with demographic variables
 - D) has not been investigated
38. One of the theoretical models behind the Circular Mood Scale is the circumplex model which
- A) classifies all moods as being basic rather than derived
 - *B) assumes there are two independent dimensions of mood
 - C) sees moods as varying in intensity and generalizability
 - D) classifies moods as being circular in nature
39. The Circular Mood Scale was presented as an example of:
- A) a laboratory procedure
 - B) a highly reliable technique
 - C) a good example of a checklist
 - *D) a visual analogue scale
40. Penile Tumescence Measurement would be classified as a(n):
- *A) laboratory procedure
 - B) self-report
 - C) unobtrusive observation
 - D) clinical interview
41. In studying reliability of penile tumescence measurement, we would typically
- A) ask the subject to self-report sexual arousal
 - B) compare sex offenders with control subjects
 - C) disregard internal consistency as inappropriate
 - *D) use film clips or slides, with some of them repeated
42. One problem with using arrest records to validate penile tumescence measurement is
- *A) the large number of false negatives
 - B) the small number of false positives
 - C) the small number of false negatives
 - D) the large number of false positives
43. In Freund's (1965) study of pedophiles and alcoholics
- A) penile tumescence measurement was too unreliable
 - *B) the results show that child molesters and controls can be differentiated
 - C) alcoholics showed the greatest arousal to female slides
 - D) many pedophiles were misclassified as alcoholics
44. The difficulty with using penile tumescence measurement to distinguish between various diagnostic groups like rapists versus child molesters:
- *A) sex offenders are usually multiple offenders
 - B) rapists do not show a sexual response unless there is violence
 - C) we do not have a theoretical framework to understand results
 - D) it is easier to arrest child molesters than rapists
45. In the study of anxiety, self report is relevant because
- A) most anxiety scales are self report instruments
 - B) motor disturbances can best be observed by others
 - C) anxiety is diffuse rather than focused
 - *D) subjective feelings of distress are central to anxiety

46. The Pleasant Event Schedule is designed to assess
- *A) the amount of external positive reinforcement a person receives
 - B) both state and trait anxiety
 - C) the effect of environmental stress upon behavior
 - D) the degree to which a person exhibits symptoms of depression
47. In taking the Pleasant Event Schedule the client indicates
- A) which events occurred, and how happy the person felt during the event
 - B) which events he/she engaged in, and why
 - *C) which events have occurred, and how pleasant each was
 - D) the client actually makes up his/her own list of events
48. The reliability of the Pleasant Event Schedule
- *A) is good, particularly the internal consistency
 - B) is rather low as expected, since events occur infrequently
 - C) is much better for college students than adult volunteers
 - D) seems to decrease with the age of the subject
49. Most of the validity evidence presented by Lewinsohn on the Pleasant Event Schedule would be considered
- *A) construct validity
 - B) inadequate
 - C) social validity
 - D) marginal at best
50. In the Bouman and Luteijn (1986) study of Pleasant Event Schedule in Dutch psychiatric patients, scores on a subset of items correlated significantly with measures of anxiety
- A) showing that the Pleasant Event Schedule is a measure of neuroticism
 - *B) when depression was statistically removed, such correlations became insignificant
 - C) casting into doubt the validity of this instrument as a measure of depression
 - D) contrary to the results obtained in U.S. samples
51. One of the problems with the Pleasant Event Schedule from a psychometric point of view:
- A) test retest reliability is low and other types of reliability cannot be assessed
 - *B) most reliability and validity data is in unpublished manuscripts
 - C) subjects are allowed to choose as many items as they wish
 - D) the focus is on behavioral issues related to depression
52. Social skills usually have two components:
- A) cognitive and affective
 - B) intensity and specificity
 - *C) interpersonal and evaluative
 - D) introversion and extraversion
53. The Rathus Assertiveness Schedule is made up of items
- A) that are vignettes where the client needs to make an assertive response
 - B) chosen because they distinguish assertiveness from aggression
 - C) that initially showed significant gender differences
 - *D) to be rated from very characteristic to very uncharacteristic
54. The validity of the Rathus Assertiveness Schedule was originally established by comparing RAS scores to:
- *A) semantic differential ratings
 - B) therapists' ratings of assertiveness
 - C) MMPI clinical scales
 - D) self ratings of aggressiveness

55. The Rathus Assertiveness Schedule is said to differentiate between assertive and nonassertive college students. This would be an example of _____ validity.
- A) social
 - B) tertiary
 - C) convergent
 - *D) discriminant validity
56. The authors of the Adult Self-expression Scale used a two dimensional table of specifications. The two dimensions were:
- A) assertiveness and aggression
 - B) positive and negative responses
 - *C) situations and behaviors
 - D) anxiety and depression
57. Most of the items in the Adult Self-expression Scale
- A) were written by the authors
 - *B) came from other scales
 - C) consist of interpersonal situations
 - D) are similar to MMPI type items
58. The items that were retained for the Adult Self-expression Scale were retained on the basis of:
- *A) item analysis
 - B) factor analysis
 - C) internal consistency
 - D) test-retest reliability
59. High scorers on the Adult Self-expression Scale describe themselves as:
- A) having great difficulties in interpersonal situations
 - B) aggressive particularly in such situations as automobile driving
 - C) more conventional, a follower rather than a leader
 - *D) more self-confident and spontaneous, more achievement oriented
60. The content validity of the items for the Children's Assertive Behavior Scale was initially established by
- A) comparison with adult scales of assertiveness
 - B) content analysis
 - *C) judgment of graduate students
 - D) factor analysis
61. In scoring the Children's Assertive Behavior Scale
- A) aggressive responses are given negative weights
 - *B) assertive responses are given a weight of zero
 - C) passive responses are given positive weights
 - D) assertive responses are given positive weights
62. One major line of evidence for the validity of assertiveness questionnaires:
- A) gender differences in both assertiveness and aggression
 - B) pattern of correlations with MMPI scales
 - C) negative loadings on social desirability
 - *D) sensitivity to changes due to therapeutic interventions
63. A major problem with many assertiveness scales:
- A) do not control for various response sets
 - B) are constructed without a theoretical framework
 - *C) they confound assertiveness and aggression
 - D) focus too much on specific behaviors

64. Originally the Fear Survey Schedule was designed to
- A) assess how phobic a patient might be
 - B) measure the degree of neuroticism present
 - C) assess phobias as a theoretical construct
 - *D) provide the therapist with an overview of a client's fears
65. The Fear Survey Schedule is somewhat unusual in that:
- *A) there are various forms developed by different authors
 - B) each revision of the scale made it longer rather than shorter
 - C) there is almost no information on its reliability and validity
 - D) it came out of a psychoanalytic framework
66. Tasto (1977) suggested that there were four major factors in the Fear Survey Schedule. Which of these is NOT one of them?
- A) fears of interpersonal events
 - B) fears related to small animals
 - *C) fears related to snakes
 - D) fears related to death and pain
67. In validating a behavioral assessment procedure like the Fear Survey Schedule, we would expect correlations with MMPI scales to be
- A) very low
 - *B) in the mid range
 - C) very high
 - D) high for patients but low for controls
68. From a behavioral point of view, validity is best established by comparing
- A) scores on an instrument like the FSS with MMPI scales
 - B) verbal reports of the client with therapists' ratings
 - C) total scores on the FSS with self ratings
 - *D) verbal reports to overt behavior and/or physiological indices
69. In attempting to validate a measure like the Fear Survey Schedule, the results from the literature suggest that:
- A) measures of latency seem valid, but measures of distance are not
 - B) such measures really have marginal validity at best
 - *C) there may well be confounding factors present
 - D) such measures are indeed quite valid
70. To evaluate a scale like the Fear Survey Schedule we can use a:
- *A) behavioral avoidance task
 - B) construct validity analysis
 - C) bimodal assertiveness technique
 - D) multitrait-multimethod design
71. The literature indicates _____ correlations between FSS scores and various behavioral indices.
- A) high
 - *B) low
 - C) median
 - D) very significant
72. Normative data on the FSS is given separately by gender because:
- A) women are substantially more fearful than men
 - B) the Test Guidelines require such information be given
 - C) women do more poorly on behavioral avoidance tasks
 - *D) women rate themselves as more fearful than men

73. The three factors of the Fear Questionnaire obtained in a sample of Hong Kong Chinese students:
- A) were rather different than the three original subscales
 - B) did not correlate with each other
 - *C) essentially mirrored the three original subscales
 - D) reflected unique cross-cultural differences
74. The Home Observation for Measurement of the Environment Inventory attempts to assess:
- A) the physical properties of a person's home
 - *B) a child's developmental environment
 - C) how the child behaves at home versus school
 - D) the socioeconomic status of a family
75. The test-retest stability of the HOME Inventory ranged from .05 to .70. We might therefore conclude that:
- *A) such coefficients may reflect the kind of changes that occurred
 - B) the reliability is rather low for such an instrument
 - C) the HOME Inventory is measuring something else
 - D) the authors need to eliminate poor items from the inventory
76. In regards to predicting IQ, the HOME Inventory
- A) was better than the Stanford-Binet
 - B) did not do a very good job
 - C) was valid with black children but not white
 - *D) was better than socioeconomic status
77. The Classroom Environment Scale covers three areas. Which one is NOT one of these?
- A) structural aspects of the classroom
 - *B) reward of curiosity
 - C) goal orientation variables
 - D) interpersonal relationships
78. The original pool of items for the Classroom Environment Scale
- A) came from a similar scale developed years before
 - B) was constructed from factor analytic studies
 - C) was developed from teachers' diaries and playground observations
 - *D) was developed using a theoretical framework, a literature survey, and interviews
79. The final version of the Classroom Environment Scale
- A) contains two forms, one for the actual classroom and one for the preferred classroom
 - B) showed substantial correlations with the Marlowe-Crowne scale
 - *C) contains nine scales with acceptable internal consistency and low intercorrelations
 - D) did not differentiate among different classrooms
80. The Differentiation scale of the Individualized Classroom Environment Questionnaire measures:
- A) whether the teacher is involved with each student
 - *B) whether different students engage in different activities
 - C) whether a classroom differs in unique ways
 - D) the degree to which a teacher allows curiosity and exploration
81. The College Characteristics Index has its theoretical roots in:
- A) Horney's theory of neurotic needs
 - *B) Murray's need-press personality theory
 - C) Moos's theory of interpersonal relationships
 - D) Watson's behaviorism

82. On the Bruvold (1968) scale to evaluate the taste of water, a hedonic scale item might be:
- A) I would be very happy to buy this water
 - *B) I dislike this water extremely
 - C) I could never drink this water
 - D) this water has a horrible taste
83. The Environmental Preference Questionnaire attempts to measure
- A) whether a person is romantic vs realistic
 - B) which environment is more stressful - urban or rural
 - *C) persons' preferences for different environmental settings
 - D) how a person feels about urban development
84. A major "problem" with the Environmental Preference Questionnaire is that
- A) its factor structure does not support its theoretical genesis
 - B) the results do not support its construct validity
 - C) its seven scales are based on factor analysis
 - *D) it may well be a personality inventory
85. The limited evidence indicates that a broader level of analysis (like rating warmth) is _____ than a more specific level, like observing hugging.
- *A) more reliable and valid
 - B) more reliable but less valid
 - C) less reliable but more valid
 - D) less reliable and less valid
86. In regards to measurement in family assessment we can conclude that:
- A) it is a well established field with great reliability and validity
 - B) most research is well replicated and uses large samples
 - C) the multitrait-multimethod approach is widely used
 - *D) currently there are multiple problems that limit its usefulness
87. In regards to family systems rating scales, we can conclude that:
- A) most have poor reliability but acceptable validity
 - *B) their research utility and usefulness are questionable
 - C) most scales are clearly defined and anchored with specific behaviors
 - D) the degree of training of the rater is an irrelevant variable
88. Two major theoretical dimensions in the area of family assessment are:
- A) marital adjustment and satisfaction
 - *B) control and cohesion
 - C) interaction and individuation
 - D) the individual and the family
89. In regards to marital quality scales we can conclude that:
- A) most are affected by either self-deception or impression management
 - B) different measures of these constructs correlate highly with each other
 - C) their factor structure is fairly well established
 - *D) most of these scales show substantial psychometric limitations
90. Some studies have found _____ between measures of social desirability and measures of marital adjustment.
- *A) modest correlations
 - B) no relationship
 - C) substantial correlations
 - D) a statistical artifact

91. The McMaster Family Assessment Device
 - A) requires a well trained observer
 - B) is based on the hypothesis that communication is central
 - *C) is a self-report instrument based on systems theory
 - D) measures one variable of overall adjustment
92. The Family Apperception Test is criticized for:
 - A) poor construct validity
 - *B) lack of content validity
 - C) being a projective technique
 - D) lack of reliability data
93. The Family Systems Test which uses the placement of figures on a board, attempts to measure
 - *A) cohesion and hierarchy
 - B) control and cohesion
 - C) family perception and behavior
 - D) communication and affection patterns
94. Sociometric procedures generally measure
 - A) control and cohesion in a family setting
 - *B) a child's status within a peer group
 - C) personality types, like the "well behaved" child
 - D) demographic variables like socioeconomic status
95. An example of the restricted nominations procedure:
 - A) do you prefer Bobby or Timmy to play with?
 - B) on a five point scale, how much do you like Margaret?
 - *C) which three children would you want on your team?
 - D) on this list of names rate how smart each child is.
96. In regards to sociometric techniques we can conclude generally that
 - A) reliability in most cases is marginal
 - B) lack of or low reliability represents error variance
 - C) test-retest reliability increases as the time interval gets longer
 - *D) the reliability of older children is greater than that of younger children
97. Which of these statements is correct?
 - A) reliability of sociometric techniques is generally inadequate
 - B) reliability of negative nominations is greater than that of positive nominations
 - *C) peer rating procedures seem somewhat more reliable than restricted nomination procedures
 - D) test-retest reliability of sociometric techniques is quite poor
98. The items on the Environmental Check List were originally obtained from:
 - *A) open-ended written descriptions of geographical localities
 - B) the dictionary and the thesaurus
 - C) standard checklists like the ACL
 - D) experiments in which students looked at slides of various cities
99. In a study of the reliability of the Environmental Check List, two samples were used both describing the same city. This type of reliability can best be described as _____ reliability.
 - A) intrarater
 - B) test-retest
 - C) alternate forms
 - *D) interrater

100. The validity of the ECL
- A) consists primarily of criterion studies
 - B) was studied using the multitrait-multimethod approach
 - *C) was supported in three separate studies
 - D) seems marginal at best
101. The ACL was originally developed
- A) as a self-report measure in behavioral studies
 - *B) as an observer check list
 - C) as a personality inventory
 - D) as a Q sort to study psychopathology
102. The distribution of items in a Q sort is usually
- A) rectangular with an equal number of items in each pile
 - *B) symmetrical and follows the normal curve
 - C) left up to the discretion of the sorter
 - D) done in the same way as the Thurstone method
103. The best known Q set is the
- *A) California Q set
 - B) ACL Q set
 - C) Ipsative Q set
 - D) Marital Q set
104. Block's Q sort of the optimally adjusted personality was developed by:
- A) observing well adjusted persons studied at IPAR
 - B) having well adjusted individuals do self sorts
 - C) Block himself through his clinical judgment
 - *D) using the modal sorting by nine clinical psychologists
105. The number of items in a Q set should be about 80 to 100. Anything longer and
- *A) the task would be burdensome for a sorter
 - B) a sorter could not keep track of all the items
 - C) the reliability would quickly become unacceptable
 - D) the total score would be too large as a sum
106. Most Q sets are unstructured. This means that:
- A) the sorter can use any distribution he/she wishes
 - B) the directions are not standardized
 - C) it is up to the sorter to decide how many piles to generate
 - *D) there was no blueprint by which to generate/select items
107. Metzger (1979) developed a Q set to assess Kubler-Ross' stages of dying. She concluded that:
- A) her results supported Kubler-Ross' theory
 - *B) Q methodology was useful to investigate terminal illness
 - C) issues of reliability were rather problematic
 - D) unstructured Q sorts were more valid than structured ones

19 The History of Psychological Testing

1. Pinel is said to represent a turning point from
 - A) philosophy to psychology
 - B) seeing mental illness as a criminal activity
 - C) an idiographic approach to a nomothetic one
 - *D) demonology to natural causation
2. Jean Esquirol is best remembered for:
 - A) allowing the patients of the Bicetre to be freed
 - *B) writing a basic text of psychopathology
 - C) proposing that hypnotism and hysteria were related
 - D) developing basic statistical procedures
3. Esquirol proposed that various degrees of mental deficiency could be distinguished by
 - A) physical signs of the patient's face
 - B) performance on a test like the Binet-Simon
 - *C) the patient's use of language
 - D) causative agents like German measles
4. Many of the procedures that _____ developed were later incorporated into tests of intelligence.
 - *A) Seguin
 - B) Esquirol
 - C) Pinel
 - D) Janet
5. In general, French investigators
 - A) emphasized both observation and statistical analysis
 - *B) focused on clinical observations but not quantification
 - C) emphasized the use of psychological tests like the Binet-Simon
 - D) focused on medical rather than psychological conditions
6. Established the first French laboratory:
 - A) Pinel
 - B) Esquirol
 - C) Janet
 - *D) Binet
7. Binet criticized the testing procedures developed by German and British investigators in that they were:
 - *A) too sensory and too simple
 - B) placing too much emphasis on imagination
 - C) too empirical and not related to theory
 - D) unreliable and too experimental

8. The Binet-Simon scale is considered to be
 - A) the first test to measure a mental faculty
 - B) the first intelligence test based on a theoretical framework
 - *C) the first standardized test of general intelligence
 - D) the first test meant to be administered by a psychologist
9. The items in the 1905 Binet-Simon scale were selected because
 - *A) they were empirically related to a number of variables
 - B) of Binet's theory about mental faculties
 - C) they were already being used in the school system
 - D) Binet believed that they were valid items
10. It's generally held that psychology as a science was "born" in the year
 - A) 1692
 - *B) 1879
 - C) 1908
 - D) 1942
11. The work of _____ marks a turning point from speculation to _____.
 - A) Janet; humanitarianism
 - B) Ebbinghaus; insight
 - C) Spranger; natural causation
 - *D) Wundt; empiricism
12. Made extensive use of the free-association method:
 - *A) Kraepelin
 - B) Ebbinghaus
 - C) Wundt
 - D) Spranger
13. Well known for his investigations of memory:
 - A) Binet
 - *B) Ebbinghaus
 - C) James
 - D) Kraepelin
14. The Allport-Vernon-Lindzey Study of Values
 - A) was developed in Wundt's laboratory
 - B) originally came from the work of Munsterberg
 - C) is a good example of the German nomothetic approach
 - *D) was widely used in social psychology
15. In Spranger's theory, the _____ value focuses on love of people.
 - A) political
 - *B) social
 - C) extraverted
 - D) religious
16. The word nomothetic can best be defined as:
 - A) focusing on individual differences
 - B) emphasizing visual perception
 - *C) designed to discover general laws
 - D) suspicious of speculation and intuition
17. Perhaps the major historical figure in the testing movement was:
 - *A) Darwin's cousin
 - B) a pupil of Wundt
 - C) a major investigator of memory
 - D) the author of the two factor theory

The History of Psychological Testing

18. _____ contributed the concept of test reliability.
 - *A) Spearman
 - B) Galton
 - C) Thompson
 - D) Pearson
19. _____ considered a test score to be the sum of truth and error.
 - A) Wundt
 - *B) Spearman
 - C) Binet
 - D) Galton
20. The so-called Scottish Surveys
 - A) compared Scottish versus French children on the Stanford-Binet
 - B) were an attempt to validate the Binet-Simon scales
 - *C) were nationwide testing of the intelligence of 11 year olds
 - D) assessed the mental health of school children
21. One of Wundt's assistants who was interested in individual differences:
 - *A) James McKeen Cattell
 - B) Joseph Jastrow
 - C) Karl Pearson
 - D) Hugo Munsterberg
22. Chronologically, which occurred first?
 - A) establishment of Wundt's laboratory
 - B) the publication of the Binet-Simon scale
 - C) Cattell's paper on mental tests
 - *D) Pinel at the Bicetre Hospital
23. Which individuals "belong" together as far as their contributions?
 - *A) Fisher and Pearson
 - B) Wundt and Pinel
 - C) Charcot and Thompson
 - D) Spranger and Darwin
24. Joseph Jastrow is best remembered for:
 - *A) demonstrating tests at the 1893 Columbian exposition
 - B) establishing the first psychology clinic in the U.S.
 - C) publishing a paper on mental tests
 - D) attempting to identify geniuses
25. Horace Mann
 - *A) replaced oral with written exams in the Boston area
 - B) made psychological tests popular at an exposition
 - C) published the first textbook on educational measurement
 - D) developed a theory based on values
26. Perhaps the major contribution of German psychologists to testing was:
 - A) the focus on the importance of sensory modalities
 - *B) the emphasis on standardization and control
 - C) the use of sophisticated statistical analyses
 - D) the need for clearly spelled out theories of behavior
27. The major focus of British investigators was:
 - A) clinical observation
 - *B) individual differences
 - C) personality assessment
 - D) educational measurement

28. Edward Lee Thorndike was affiliated with
 - A) the Vineland Training School
 - B) the University of Chicago
 - *C) Columbia University
 - D) Clark University
29. Sharp (1899) administered a battery of mental tests to a sample of students. Today she would be severely criticized because:
 - *A) her sample was very small
 - B) her test items were taken from the Binet-Simon
 - C) she was not concerned with reliability
 - D) she did not use the multitrait-multimethod approach
30. Clark Wissler's (1901) study
 - A) looked at the test-retest reliability of measures of imagination
 - B) found that grades in different courses did not correlate
 - C) was meant to be a severe criticism of early testing
 - *D) compared Cattell's tests with anthropometric measures and grades
31. When the 1905 Binet-Simon scale was published
 - A) the scale was accepted as evidence that Sharp's study was wrong
 - *B) American university psychologists generally showed little interest
 - C) a large number of American university psychologists translated it
 - D) it was used to restrict immigration to the United States
32. Which two would agree that intelligence is a genetically fixed quantity?
 - *A) Goddard and Terman
 - B) Galton and Binet
 - C) Wundt and Munsterberg
 - D) Hall and James
33. The study of the Kallikak family
 - A) was presented to further the point that IQ was inherited
 - B) was undertaken by Galton to study evolution
 - C) was the first to look at reproduction in the mentally retarded
 - *D) was taken as evidence of the hereditary nature of pathology
34. His work made the Binet test the best known intelligence test in the United States.
 - A) Stanford
 - *B) Terman
 - C) Goddard
 - D) Wechsler
35. One of the major innovations of the 1916 Stanford-Binet:
 - A) using census data
 - B) having standardized instructions
 - C) focusing on the process rather than the result
 - *D) the use of IQ
36. The 1916 Stanford-Binet
 - A) was used exclusively to identify gifted children
 - B) was basically a translation of the 1911 Binet-Simon
 - *C) made the public accepting of testing
 - D) originally was intended to identify mentally retarded children
37. Which of these was designed to be used with the blind?
 - *A) Hayes-Binet
 - B) Yerkes Point Scale
 - C) Kuhlmann-Binet
 - D) Whipple's Manual

The History of Psychological Testing

- 38. A major figure in the early growth of American psychology:
 - A) Lewis Terman
 - *B) Granville Stanley Hall
 - C) James McKeen Cattell
 - D) William Stern
- 39. The first president of the American Psychological Association:
 - A) Henry Goddard
 - *B) Granville Stanley Hall
 - C) Lewis Terman
 - D) William James
- 40. Which of these does NOT belong:
 - A) introduced the psychoanalytic movement to the U.S.
 - B) was a central figure in establishing child psychology
 - *C) authored the study of the Kallikak family
 - D) established several psychological journals
- 41. What was unique about the Healy-Fernald series of tests?
 - A) included only non-verbal items like puzzles
 - *B) emphasized the process rather than the outcome
 - C) were extremely well standardized
 - D) represented the first projective technique other than the Rorschach
- 42. The advantage of the Army Alpha and Beta over the Stanford-Binet:
 - *A) could be group administered
 - B) yielded scores on Verbal and on Performance IQ
 - C) could be self-administered by the subject
 - D) alternate form reliability could be assessed
- 43. The Woodworth Personal Data Sheet
 - A) was designed to identify candidates for officer training
 - B) was a replacement for more costly psychiatric interviews
 - *C) was essentially a primitive version of the MMPI
 - D) was a good example of biodata questionnaires
- 44. A major criticism of the Woodworth Personal Data Sheet:
 - *A) did not address issues of reliability and validity
 - B) it was a paper-and-pencil version of a psychiatric interview
 - C) it was longer than the MMPI
 - D) it was not particularly successful
- 45. After World War I, testing
 - *A) became an integral part of the U.S. culture
 - B) was used only in educational settings
 - C) was of little interest because of Cattell's work
 - D) had negative connotations because of the military experience
- 46. A test that quickly became a favorite cross-cultural tool:
 - *A) draw-a-man
 - B) Rorschach
 - C) Stanford-Binet
 - D) Downey Will Temperament
- 47. In the field of educational testing it was the _____ that made large scale testing a practical reality.
 - A) 1916 Stanford-Binet
 - B) Army Alpha
 - C) Healy-Fernald tests
 - *D) Stanford Achievement Test

48. The Downey Will Temperament Test is mentioned as an example of:
- A) a test that revolutionized educational measurement
 - *B) a personality test that did not work
 - C) an opposite approach to that taken by the Stanford-Binet
 - D) one of the few successful tests from Wundt's laboratory
49. The name of Louis Thurstone is associated with:
- *A) primary mental abilities
 - B) two-factor theory
 - C) genetic studies of genius
 - D) Berkeley Growth study
50. A pioneer in the field of infant testing:
- A) Florence Goodenough
 - *B) Arnold Gesell
 - C) Louis Thurstone
 - D) Horace Mann
51. The Scholastic Aptitude Test was introduced in
- A) 1906
 - *B) 1926
 - C) 1946
 - D) 1966
52. Which two are correctly paired?
- *A) Buros; Mental Measurements Yearbook
 - B) Hartshorne and May; TAT
 - C) Morgan and Murray; Rorschach
 - D) Goddard; genetic studies of genius
53. The Stanford-Iowa controversy centered over
- A) the definition of coaching
 - B) whether scores on the SAT could be improved
 - C) whether perception reflected pathology
 - *D) the constancy of the IQ
54. Wechsler argued that:
- A) intelligence was not hereditary but influenced by nurture
 - B) a primary mental abilities approach was the most valid
 - C) intelligence tests should focus on normal children rather than geniuses
 - *D) available adult tests of intelligence had been developed from children's tests
55. The Army Alpha of the first World War was replaced by the _____ in the Second World War.
- A) Wechsler Adult Intelligence Scale
 - B) Bender Visual Motor Gestalt Test
 - *C) Army General Classification Test
 - D) Diagnostic Psychological Test
56. Perhaps the best known test associated with the World War II period is the:
- *A) MMPI
 - B) Stanford-Binet
 - C) Wechsler-Bellevue
 - D) Army General Classification Test
57. The Authoritarian Personality:
- A) developed a unified theory of the human intellect
 - B) was a direct outgrowth of the work of Hathaway at Minnesota
 - *C) attempted to understand psychoanalytically the Nazi movement
 - D) severely criticized American efforts at psychological testing

The History of Psychological Testing

58. Which occurred first chronologically?
- A) the Stanford-Iowa controversy
 - *B) the first Mental Measurements Yearbook
 - C) Jensen and the IQ controversy
 - D) Guilford's structure of intellect model
59. The textbook argues that the launching of the Soviet Sputnik
- *A) fostered large scale testing with multiple choice exams
 - B) led to controversies over the use of the MMPI in the Peace Corps
 - C) once again brought the nature-nurture issue to the fore
 - D) resulted in truth-in-testing legislation
60. Banesh Hoffman's book (The Tyranny of Testing) argued that:
- A) tests like the Stanford-Binet were clearly biased
 - *B) multiple choice exams penalized creative students
 - C) college entrance should be open to all
 - D) national scholarship programs were communist inspired
61. The anti-test movement of the 1960's
- A) dealt a severe blow to testing
 - B) really had little effect on psychological testing
 - *C) made psychologists more aware of issues like privacy
 - D) focused on racial bias
62. The Bell Curve is historically related to
- A) Healy and Fernald
 - *B) Goddard and Terman
 - C) Thurstone and Gesell
 - D) Bayley and Goodenough
63. The concern over the decline of SAT scores
- A) resulted in truth-in-testing legislation
 - B) was again interpreted in nature-vs-nurture terms
 - *C) brought testing once again to the public forum
 - D) culminated in the book the Bell Curve
64. Minimum competency testing primarily concerns
- A) entrance into college
 - B) the content validity of achievement tests
 - C) the nature-nurture controversy
 - *D) high school graduation
65. The textbook predicts that in the future tests will
- *A) still be with us
 - B) be more sensory
 - C) not be used as widely
 - D) be more essay than multiple-choice