Computerized Neuropsychological Assessment: The Good, the Bad, and the Ugly

Mike R. Schoenberg, Ph.D., ABPP-CN
Department of Neurology
University Hospitals Case Medical Center

Disclosures

- None

Objectives

- Neuropsychological Evaluation
- Bases of Assessment
- "Traditional" Neuropsychological Assessment
  - Measures
  - Benefits/Strengths (the good)
  - Problems/Weakness (the bad and ugly)
- Computerized Cognitive Assessment
  - Measures
  - Benefits (The good)
  - Problems (The bad and ugly)
- Conclusions
- Future Directions

Neuropsychological Evaluation: Fundamentals

- Study of Brain-Behavior Relationships
  - Identifies presence (or absence) of neuropsychological Deficits
- Assumptions for Evaluation
  - Brain dysfunction affects behavior
  - Behavior changes can be associated with particular brain processes/areas/neurological syndromes
  - Assessment can be reliable
  - Assessment can be valid
  - Assessment affects diagnosis/treatment

Neuropsychological Evaluation: Traditional

- Assessment versus Evaluation
  - Assessment
    - Collection of historical data
    - Collection of cognitive data
    - Collection of mood data
  - Evaluation
    - Interpretation of data for diagnosis/treatment planning

Purpose of Assessment

- Screen for presence/absence of potential problem or change
  - Is performance above or below threshold to identify possible problem?
    - Common example is MMSE or Clock Drawing task
  - Screening data, in and of itself, typically not diagnostic nor used for treatment planning
- Diagnosis of problem (etiology) and plan treatment
  - Is data suggestive of known syndromes/diagnostic entities?
  - Interpretation of data for diagnosis/treatment planning

Neuropsychological Evaluation: Assessment Methods

- Historical information
  - Referral question(s)
  - Presenting problems
  - Historical information
    - Other laboratory tests
    - Comorbid conditions
    - Other historical data
- Clinical Interview
- Behavioral observations
  - Neurobehavioral tests/sensory/perceptual/cranial nerves/etc.
- Neuropsychological test administration
  - Paper and pencil based cognitive tests
  - Computer assisted tests
- Psychological test administration
- Assessment of effort/Task engagement

Assessment: The Basics

- Assessment measures must be:
  - Reliable
  - Valid
  - Efficient
  - Sensitive (and specific)

Assessment: The Basics

- Reliability:
  - Reliability typically refers to consistency in measuring a construct
    - A test is only as valid as it is reliable
  - Reliability includes internal consistency, test-retest reliability, alternate forms, etc.
    - Internal consistent is how consistent items within a test are to measuring the construct
    - Test-retest reliability how consistent across time. Assumes construct does not change over time.

Assessment: The Basics

- Validity
  - Extent test measures what it intends to measure
    - Criterion Related validity
      - Predictive (criterion) Validity
        - High score predict behavior/deficit (dementia)
      - Concurrent validity
        - Does shorter test measure same construct as validated longer test?
    - Construct related validity
      - Convergent/discriminant validity
        - do scores differ between groups with syndromes/dx in which test scores should theoretically differ
    - Ecological validity
      - Extent test predicts a “real world” behavior or problem thought to be associated with construct (e.g., good driving)

Sources of Assessment Error

- Measurement error
  - Assessment a "picture in time"
    - Variation in CNS pathology
  - Test not perfectly valid
    - Some error in tests
  - Sampling error
    - Selection of tests and test items
  - Scoring/Administration errors
    - Intra-rater reliability
    - Inter-rater reliability
  - Patient variables
    - Task engagement/motivation to perform well
    - Educational/occupational/cultural/language/age factors
- Test score = syndrome + measurement error + premorbid ability + drugs + effort + practice

Diagnostic Characteristics

	Test YES	Test NO	
Disease: YES	(a) True +	(b) False –	SENS
Disease: NO	(c) False +	(d) True –	SPEC
	PPP	NPP	Hit Rate

Sensitivity: a/[a + b]. It's there, and you see it

Specificity: d/[c + d]. It's not there, and you don't see it

Positive Predictive Value: a/[a+c]. Your test says it's there, and it is

Negative Predictive Value: d/[b+d]. Your test says it's not there, and it's not

Neuropsychological Evaluation

- Important shift in Neuropsychological Assessment
  - NOT can a cognitive test discriminate abnormal from healthy?
  - Rather, can cognitive test/battery discriminate subtypes of diseases or phases of single disease.
  - MMSE is highly sensitive, not specific
    - That is, if you score low on a test, suggestive something is wrong, but don't know what.
- Diagnostic Characteristics dependent upon prevalence of disorder

Diagnostic Characteristics

N = 100. Prevalence rate = 11%

	Test YES	Test NO	
Disease: YES	10 (a)	1 (b)	11
Disease: NO	14 (c)	75 (d)	89
	24	76	Hit Rate

Sensitivity (SENS): a/[a + b] = 91%

Specificity (SPEC): d/[c + d] = 84%

Positive Predictive Value (PPV): a/[a+c] = 42%

Negative Predictive Value (PPV): d/[b+d] = 99%

Hit Rate (HR) = 85%

Diagnostic Characteristics

N = 100. Prevalence rate = 50%

	Test YES	Test NO	
Disease: YES	46 (a)	4 (b)	50
Disease: NO	8 (c)	42 (d)	50
	54	44	Hit Rate

Sensitivity: a/[a + b] =92%

Specificity: d/[c + d] =84%

PPV: a/[a+c] = 85%

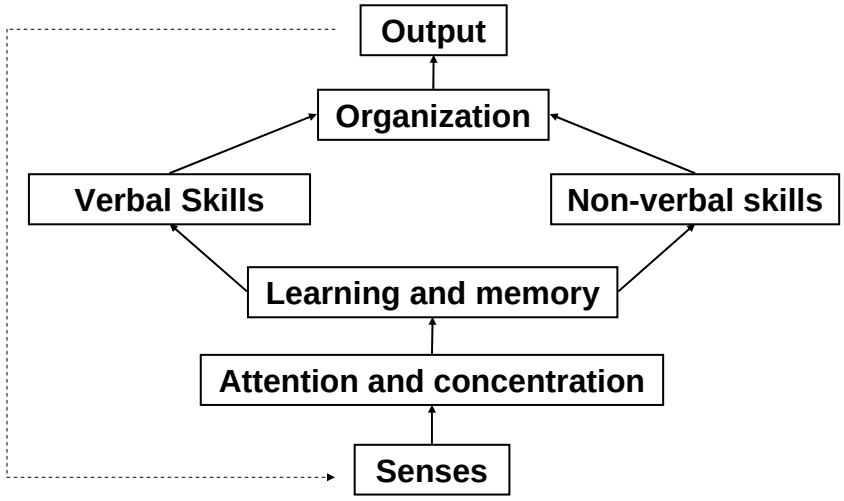
NPV: d/[b+d]. 95%

Hite Rate = 90 %

Variables that effect diagnostic test characteristics

- Prevalence rate affects PPV and NPV.
  - Screening for a syndrome better with higher prevalence rate
    - Positive Predictive Value (power) of test increases with higher prevalence
- Need to balance adverse affect of making false positive error versus a false negative error
  - Screening ideal when consequence of false positive is low while consequence of making a false negative error is bad

Brain Function Organization



Baker GA. *Personal Communication, 2008*

Neuropsychological Evaluation: Diagnostic Assessment Measures

- Measurement of cognitive constructs
  - General Cognitive Ability (IQ)
  - Achievement (academic development)
  - Processing Speed/psychomotor speed
  - Attention/Concentration
  - Memory
  - Language
  - Visuoperceptual/Visuoconstructional
  - Executive functions (problem solving, insight, judgment, etc.)
- Psychological Function

Traditional Neuropsychology Battery: Diagnosis

DOMAINS	TESTS
General Cognitive (IQ)	Intelligence Test (e.g., Wechsler Adult Intelligence tests) Achievement Tests
Psychomotor Speed	Finger Tapping, Grooved Pegboard, Continuous Performance Tasks (reaction time)
Attention/Executive Function	Trail Making Test A & B Coding Tasks (e.g., symbol digit substitution) Letter-number sequencing (working memory) Stroop color-word tasks (inhibition/interference) Wisconsin Card Sorting Test
Learning & Memory **Immediate (short-term)** **Delayed (long-term)**	Wechsler Memory Scales Auditory Verbal Learning Tasks Rey-Osterreith Complex Figure memory
Language **Receptive, Expressive, & repetition**	Boston Diagnostic Aphasia Exam Verbal Fluency Tests (semantic and phonemic) Token Test
Visuoperceptual/construction	Complex figure tasks, block design tasks
Mood	MMPI, Beck Depression Inventory

Standard Neuropsychological Battery: The Good

- Empirical support for use of Neuropsychological tests to identify brain dysfunction
  - Test Measures are Reliable
    - Inter-rater reliability
    - Intra-rater reliability
    - Test – Retest (some better than others)
    - Internal reliability
  - Test Measures have validity
    - Criterion Validity - Associated with known brain damage
      - Poor validity for some known lesions (e.g., frontal)
      - Used for diagnosis of brain dysfunction
    - Construct validity - Associated with theories of cognitive function
    - Discriminant validity – Test scores differ between diagnostic groups test scores should differ (some better than others)
    - Predictive validity - Predictor of cognitive/surgical outcome
      - Epilepsy surgery (primarily for non-lesional)
      - Deep Brain Stimulator (DBS) surgery (?)

Standard Neuropsychological Battery: Strengths

- Diagnostic value
  - Good to discriminate Neurological disorder from normal (old news)
    - Required for dx of Mild Cognitive Impairment (MCI)
      - Memory impairment = <10th %ile of normal
    - Required for HIV-associated cognitive impairment
    - Good for Dementia
    - Fair to Good for effect of medications (AEDs) on cognition
  - Fair (even good) discriminating BETWEEN neurological disorders
    - Distinct (relatively) neuropsychological profiles can distinguish
      - Alzheimer's dementia from Frontotemporal dementia vs. vascular dementia
      - Dementia from Pseudodementia
      - Parkinson's dementia from Lewy Body Dementia
- Ecological validity
  - Independent predictor of cognitive outcome (and sz freedom) from temporal lobectomy for intractable epilepsy
  - Poor ($p > 0.05$) for criterion of self-reported cognitive problems
  - Fair to Good for criterion of ADLs (67%)
  - Poor to Fair for safe driving

Standard Neuropsychological Battery: The Bad

- Benefit versus Cost?
  - Time intensive
    - Typical Assessment period ranges from 2 - 8 hours
  - Special equipment/training needed
    - Professional/training component demand is high
    - Equipment must be secured/Testing areas needed.
  - Moderate expense
    - Typical Neuropsychological evaluation cost [$2,000-3,000].
    - Typical charge for MRI [$2,500 - $4,000]
    - Typical charge for ambulatory EEG [$1,400 - $2,500]
- Limited Accessibility
  - Access to quality neuropsychological services limited
  - Because time intensive difficult to do many evaluations in short time
- Patient time intensive.
- Practice effects are present
  - Limited availability of alternate forms
- Ecological validity limited
  - poor for criterion of self-reported cognitive problems
  - Fair to poor for criterion of safe driving

Neuropsych: Diagnostic Accuracy

	SENS	SPEC	PPV	NPV	Hit Rate
DAT versus Normal					
MMSE	71-92 (82)	56-96 (98)	15-72^ (68)	64-99 (99)	(84)
Neuropsycologic	83-100 75!	84-100 74!	83-100 90!	73-100 50!	81 – 100 75!
Function Activity Q	90	90			
MRI	80-82#	87#- 90	89—90#	77#- 82	85-92#
DAT from Depressed					
MMSE	80	100	100	83	90
Neuropsychologic	83*	83*-100	86*-100	79*-90	83*-95
Pocket smell test	95	100	100	90	97.5

Note: ^10% prevalence of dementia; # 59% prevalence rate of pathologically confirmed AD (Gosche et al., 2002); *RAVLT only, 56% prevalence rate of dementia; ! = Neuropsych. testing predicting 6 year post-mortem pathologic diagnosis of DAT.

Neuropsychological: Diagnostic Accuracy

	SENS	SPEC	PPV	NPV	Hit Rate
FTD vs. Normals*					
MMSE					
Neuropsychologic			64	98	90
Computer Battery					
MRI	49-80	90	70	82	79
HIV-Dementia vs. Normal					
HIV Dementia Scale	30	0	0		
Neuropsychologic	43-100	91	83-100		

Note: *Prevalence of FTD = 16%, AD prevalence = 43%; Walker et al., 2005;

Neuropsychological: Diagnostic Accuracy

	SENS	SPEC	PPV	NPV	Hit Rate
AD vs. Vasc. Dementia					
MMSE	51	48	51	48	51
Neuropsych. Battery	75	93	-	-	89
FTD vs. AD vs. Normals*					
MMSE	-	-	-	-	-
Neuropsychologic	69	93	64	95	90
SPECT^	90	100	100	91	95
MRI volumetrics	49-80	-	-	-	-

Note: * FTD Prevalence = 16%, AD prevalence = 43%, Walker et al., 2005; ^ prevalence = 50%

Neuropsychological: Diagnostic Accuracy

	SENS	SPEC	PPP	Hit Rate
Predict Sz. Free outcome from ATL				
Left ATL	-	-	92	80
Right ATL	-	-	90	83
Adult ADHD				
Neuropsychologic		40-80	83-100	70-90

Note: *10% prevalence of dementia; Lovejoy et al., 1999

Objectives

- Neuropsychological Evaluation
- Bases of Assessment
- "Traditional" Neuropsychological Assessment
  - Measures
  - Benefits/Strengths (the good)
  - Problems/Weakness (the bad and ugly)
- Computerized Cognitive Assessment
  - Measures
  - Benefits (The good)
  - Problems (The bad and ugly)
- Conclusions
- Future Directions

Computerized Cognitive Assessment: Why

- Increasing need for fast and efficient cognitive testing
  - 58% of PCP physicians reported MMSE was too time intensive (Tangalos et al., 1996)
  - MMSE ineffective at screening unselected cases
    - Positive Predictive power = 32% (1.7% prevalence rate)
- Neuropsychological testing time intensive
- Neuropsychological resources limited
- Neuropsychological assessment difficult to apply in varied environments
  - Expensive or impractical to have trained person in environment
- Development of computer assisted administered neuropsychological/cognitive tests
- Need re-training to maintain collection of good data

Computer Cognitive Measures: Historical Development

- Historical Tests
  - Adaptation of paper/pencil neuropsychological tests to computer
    - Wisconsin Card Sorting Test
  - Development of novel computer administered tests to measure unique cognitive functions
    - Reaction time/attention tests
    - N-Back tests for working memory/application to fMRI
- Novel automated Computerized assessment
  - Department of Defense (DoD) cognitive test applications

Computer Cognitive Measures: Whom

- Department of Defense
- Federal Aviation
- Private commercial enterprises
  - Dementia
  - Concussion/sports medicine
- Psychological Test Publishers
  - Adaptations of current tests
- Pharma/National Institute of Health

Computer Cognitive Measures: What

- Current "stand alone" computerized cognitive batteries
  - Many (13) independent developers.
  - Some developers have multiple "models"

Automated Neuropsychological Assessment Metrics (ANAM)®

CNS Vital Signs ®

CANS-MCI
(Computer administered neuropsychological screen)

Cambridge Neuropsychological Test Automated Battery (CANTAB)®

CNT

CogSport®

Headminder ®

ImPACT ®

MicroCog®

NeuroTrax – Mindstreams®

Neurobehavioral Evaluation System – 3 (NES-3) ®

Specialty Automated Systems®

Computerized Cognitive Assessment: Benefits

- Easy to administer and score
- Relatively fast (.25-2 hours)
- Relatively inexpensive
- Readily available
- Applicable to many "real world" settings
  - Adaptation to PDAs or multi-function cell phones
  - Testing can occur in variety of environments
    - Battlefield, Schools, Churches, hospitals/clinics
- Can provide accurate assessment of cognitive functions
  - Information proc. speed, reaction time, working memory, etc.
- Administration procedure can be held constant
- Alternate versions can be easily developed
- Conclusion:
  - Particularly suited for screening instrument and/or monitoring change in cognitive function(s) over time

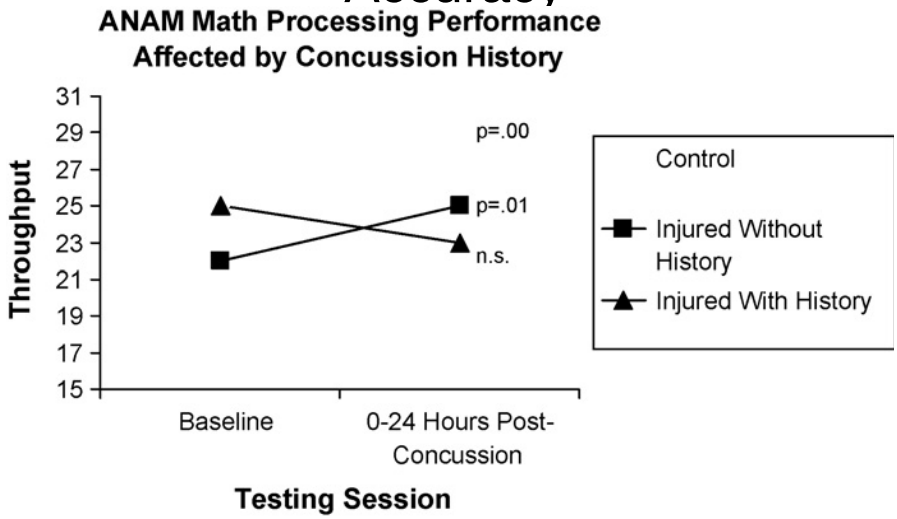
Computerized Assessment: Advantages and drawbacks

- Drawbacks
  - Reliability has been fair to poor. Much improved recently.
  - Validity is often questionable
  - Generalizability may be poor
  - High potential for abuse
    - Effect of False Positive on person/system
    - Effect of use without adequate training in neuropsychology and psychometrics
  - Few include measures of task engagement/motivation
  - With increasing sophistication comes increasing time
    - Time for some full batteries not inconsequential (1-2 hours)
  - Computer/software error
    - Losing data/Data corruption
    - Effects of different screens on reliability/validity

Neuropsychological: Diagnostic Accuracy

	SENS	SPEC	PPV	NPV	Hit Rate	ROC curve
MCI vs. Normals						
MMSE	24-72	72-97	31-55	88-94	72-86	.803
Neuropsych.	100	100	100	100	100	.89 -1.0
Computer battery	70-80	39-84	45-55	95	83	.68-.89
Concussion						
Neuropsych. Battery	88	90	-	-	89.5	-
Computer Battery	82	89	-	-	87.5	-
Brief Neurpsych.	94	76	-	-	-	-

Neuropsychological: Diagnostic Accuracy



Warden et al., 2003

Computerized Assessment: Conclusions

- Reliability is poor to good
- Validity is poor to clinically adequate (80%)
  - Detailed (independent) studies limited to military applications
  - Independent testing is otherwise generally lacking
  - Studies limited to distinguishing abnormal from healthy
- Adaptable and flexible for rapid development
  - Easy to administer and scoring is automated
  - Relative short test batteries
  - Alternate test forms readily developed
  - Can be adaptable to person during testing
- Appears best suited for:
  - Screening large numbers of individuals
  - Best for discriminating normal from abnormal
    - NOT for differential diagnosis
  - Evaluation of change in populations with risk or in treatment groups
    - Military applications
    - Application to Phase III trials
      - Evaluate for change in cognitive function with intervention
    - Sports medicine/concussion management

Neuropsychology: Extinction or Growth?

- Automated cognitive tests reflect an extension of traditional neuropsychological tests
  - Application in screening situations should "trigger" an in-depth evaluation to assess for etiology for abnormal finding
    - Dementia screening
    - Concussion management
    - Cognitive impairment in neurological/psychiatric syndromes
    - Cognitive impairment after surgical procedures
    - Cognitive impairment in other medical conditions/therapeutics
      - Lupus, Chronic Fatigue syndrome, Hep C, Chemotherapy, HIV
  - Application in monitoring for change with treatment can document + and – of medical therapies on cognition
  - Use as baseline against which to compare changes following known or suspected brain injury
- Does not replace need for trained professional in neuropsychology!

To Screen asymptomatic persons or not? Cost versus Benefit

- Automated cognitive screening may identify subtle problems before appreciated. BUT-
- No empirical evidence to support routine screening for dementia in primary care settings for asymptomatic individuals (American Academy of Neurology, Canadian Task Force on Preventive Health Care, U.S. Preventive Services Task Force).
- Impact of False Positive errors in diagnosis not established
  - Case reports of suicide in early diagnosed patients with Alzheimer's disease

Other Developments

- "Traditional" neuropsychological evaluations are
  - Getting shorter
  - New tests developed to better measure
    - Frontal lobe dysfunction
    - Non-verbal Memory
  - Combining neuropsychological measures with neurophysiological assessment (SAM).
  - Developing "screening" paper and pencil based neuropsychological batteries for specific populations
    - Alzheimer's dementia (CERAD)
    - Multiple Sclerosis
    - Parkinson's disease
    - Epilepsy (EpiTrax)
    - Schizophrenia (MATRICS)
    - HIV – dementia
    - Sports Concussion (SAC)
      - $22-27 per athlete versus $669-677 per athlete for computer

Neurophysiological Measures

- EEG and Event-Related Potentials (ERPs)
  - More sensitive to medications (and sleep deprivation) than neuropsychological measures
- Sustained Attention and Memory (SAM) test
  - Combined assessment of cognitive performance with EEG/ERPs
    - Computerized cognitive battery with working memory and episodic memory tasks.
    - Working memory assessed with spatial n-back tasks
      - Easy task, which is a 1-back test (compare spatial location of a stimulus to location of previous trial)
      - Harder task, a 2-back task.
    - Episodic memory task (24 word memory recognition task)
    - EEG recording during cognitive tasks

Neurophysiological Measures

- Measures (dependent variables)
  - Cognitive errors
  - EEG power in delta and theta ranges along with peak alpha
  - ERPs
    - Working Memory - parietal P300 peak
    - Word recognition – frontal slow wave
- Cost versus benefit?
  - Administration Time is $\leq$ 1 hour
  - Specialized equipment needed
  - Training and equipment relatively expensive
- Sensitivity and specificity excellent
  - Sensitivity 75 – 100%
- Ecological validity
  - ?

Neuropsychological vs. Neurophysiological (ERPs) Measure

	SENS	SPEC	ROC Area under curve
CBZ vs. LEV			
SAM exam	96	100	1.000
Neuropsychologic	75	75	0.888
CBZ vs. No-drug			
SAM exam	100	100	1.000
Neuropsychologic	89	96	0.997
LEV vs. No-drug			
SAM exam	75	93	0.858
Neuropsychologic	46	82	0.675

Meador et al., 2007

Drawbacks to Neurophysiologic and Neuropsychologic

- Neurophysiological assessment
  - Training and equipment costs higher than neuropsychological
  - Very high professional technical component
  - Administration time can be shorter than neuropsychological
- Increase in patient discomfort
  - Requires EEG placement on participant

Bottom Line

- Automated Cognitive Batteries here to stay
- Many options
- Little reliability and validity data to date
- Application possibilities is large
  - Screening purposes
  - Application to clinical trials (Phase III)
- Complements “traditional” neuropsychological measures
- Because dx of cognitive disorder is a medical dx, it has legal implications
  - Document methods for determing dx meticulously.
    - Other possibilities (neurological/medical) must be ruled-out
    - Should NOT be based on automated test alone

Bottom Line

- High potential for abuse, but some clear advantages
  - Enables rapid, efficient, and cost-effective screening for cognitive disorder
- Interpretation of results must be made by qualified physician or psychologist
  - Knowledge of psychometrics
    - Reliability and Validity of Measure. Sources of Error.
  - Knowledge of neurological/psychiatric syndromes
  - Cause of abnormal result may not be brain dysfunction
- Limitations of billing and practice implications
  - Doctors limited to use professional neuropsychological procedure codes
    - Most document time
  - Automated computer based neuropsychological testing procedure code available.
    - Must document time
    - Reimbursement lower than professional procedure codes

Questions

Diagnostic Value: Shifting Priorities

- Old priorities
  - Test needed to distinguish abnormal from healthy
    - Experimental designs compared individuals with known diagnosis to healthy individuals
  - Test need to be valid for criterion (brain injury)
    - Does test localize/lateralize brain dysfunction?
- New priorities
  - Test need to distinguish between abnormal syndromes (e.g., distinguish DAT from VaD from LBD)
  - Test need to have ecological validity
  - Test need to be efficient and cost effective
  - Need to affect patient outcome

Neuropsychological Evaluation: Diagnostic Evaluation

- Assessment for Purpose of diagnosis and plan treatment
  - Interpretation of data for diagnosis/treatment planning
  - Is data suggestive of known syndromes/diagnostic entities?

Neuropsychological Evaluation: The Evaluation

- Methods of Evaluation
  - Integration and interpretation of historical data, observations, and neuropsychological test scores based on the:
    - Association with known syndromes (diagnosis), and
    - Association with functional neuroanatomy, and from this:
      - Associate conclusions with interventions