

A Primer of Understanding Psychological Tests

Robert Griffith, Psy.D.
December 7, 2010

Agenda

- Difference Between Psychics and Psychologists
- Measurement in science and psychology
- A little bit about how psychological tests are developed
- Types of tests scores and tests
- Reasons testing can be helpful with examples
- When testing may be misused
- Difference between testing and psychological assessment
- Elements of a psychological report: keeping it real
- When to refer for an assessment
- How to make a referral to an astrologer if psychological testing is unhelpful to your client

What is a psychologist

- According to the American Psychological Association, those with doctorates in the field psychology are psychologists
- Doctorate degree: Ph.D., Ed.D., Psy.D.
- According to Michigan law, a limited licensed psychologist (MA, MS) may refer to his/herself as a psychologist who is supervised by the above
- Requirements of supervision vary, depending on work setting
- Neither the APA or the licensing board of psychology require continuing education credits for psychologists

Psychologist

- For LLP and LP levels, only one psychological testing course is required, (hopefully) a prerequisite class in test and measurement
- These courses may be IQ testing (most usually limited to studying the Wechsler Scales) or b) a course on personality testing (most usually limited to studying the Minnesota Multiphasic Personality Inventory)
- Practicing psychologist spend about 10 the 25% of their time connecting psychological assessments: there has been a general decrease over the last several years, regarding the time, psychologists been performing assessments
- In the 1940s and 1950s, testing was synonymous with being a practicing psychologist: psychologist now spend time teaching, administration, consultation, and direct treatment
- Assessment psychology is an area of psychology devoted to the understanding people's behavior through the use of psychological testing or other techniques

Criterion Referenced Testing

- These use cut scores, so the interpretation is usually a "win-lose" situation
- You probably remember these from school or an important exam, but are less likely to remember them if you passed them than if you failed them!
- They may assess a domain of knowledge, to determine master over a subject
- An exam to measure a law enforcement officer trainee's knowledge of laws
- EPPP...The psychology licensing exam
- NCE....Counseling exam

Psychics



- Psychics are paid people who use tarot cards, tea leaves, the I Ching, telepathic skills, crystal balls, astrology charts, Runes, Feng Shui, Kung Fu, Tofu, Palmistry, Ouija Boards, Numerology, Channeling, Remote Viewing, Vulcan Mind Meld
- They have a number of assessment tools, but do not have a standardized procedure or scoring method and are not licensed professionals
- As far as I know, they do not have a criterion referenced exam to pass (unless you count the driver's license exam)

Brief History of testing

- Tests-as pre-employment measures (of ability, skill, personality) were first given in ancient China
- Emperor would screen for fitness for duty by
 - Inferred intelligence from interviews
 - Observed traits across situations
 - Work sample of skills needed

Measurement

- Usually, texts on testing skip from Ancient China to the 19th Century and discuss Galton, Darwin's cousin
- Much going on in science in that interim period however and it generally had to do with measurement, developing the scientific method in the West
- Science, as we know it know, is all about close observation, measurement, hypothesis testing, critical thinking, weighing out alternative explanations...

Measurement

- As instruments were improved, more precise measurements were possible
- Data was always there; the instrument changed: scientists were able to measure distances, speed, temperature, weight, etc, with greater accuracy
- Eventually, scientists could observe/measure things that were not visible
- Psychology began measuring "latent" variables and constructs in the early part of the 1900s with IQ and personality tests

Measurement

- Decisions in many fields are based on measurement
- The "hard sciences" like physics, chemistry are obvious examples
- All of the social sciences have a measurement component, which may include simple or complex coded observations, surveys/self report
- Behavior is sometimes straightforward to measure, sometimes much more complex

Variables

- An important idea in psychological testing is the measurement of a latent variable(s)
- In psych testing, a variable is any characteristic of a person
- Usually, we do not use discrete variables in testing (these are simply counted or added things/events)
- In psychology, we use continuous variables. Variables are expressions of constructs
- Constructs are things like intelligence and personality

Distribution of Variables

- A psych test takes a sample of a person's behavior, to measure a construct
- The test score the person achieves is located in a distribution of scores
- This distribution of scores is also a sample, but a sample from a population
- The sample is usually of males/females, sometimes combined gender groups, stratified age groups

Variables, Samples and Normative curve

- Variables from a sample are generally distributed along a bell shaped curve
- We can make inferences, as long as we have a large enough sample
- Once we know the mean (average), we can calculate various points from the mean (called z scores) then plot standard scores
- Just about everyone, on a norm referenced test, somewhere around the mean

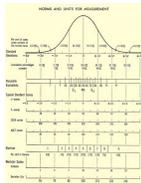
Norm referenced tests



- T scores are a type of score used on personality measures
- T scores have a mean of 50 (roughly 2/3) 68 percent of people score, for example, on the MMPI, between a score of 40 and a score of 60
- IQ scores are referred to as Composite scores or sometimes Index scores
- The smaller tests that make up IQ tests like the Wechsler Scale scores are called subtests
- Also, 2/3 of people score between 85 and 115 on IQ tests
- Percentile ranks can be helpful in explaining such scores, but should not be confused with percent correct (which would apply to criterion referenced tests)
- Percentile rank is how many people (on IQ tests, this would mean how many people in your age bracket) scored at...or below the score obtained

Curve

- Bell curve



Sample

- Samples of people are usually randomly selected from a demographic during the development of a test
- The sample for the MMPI-2 was based on the make up of the 1980 Census
- The original MMPI sample (circa 1940s) was based on a local (convenience) sample from the University of Minnesota Hospital (psychiatric patients) and "normals" or local folks that volunteered
- The Wechsler Bellevue Scale (original Wechsler IQ test) was based on a sample collected by David Wechsler in the Brooklyn New York area
- The 2008 WAIS-IV sample was based on a sample of 2200 people from the US

Reliability and Validity

- Commercial test publishers have a responsibility to follow well established guidelines for making reliable measures
- Reliability means tests are consistence measures of whatever they measure (several types of reliability, you can Google it if you are interested)
- Tests also have to be valid, that is, they have to measure what they are intended to measure

Standardized Administration

- Tests are developed and then published using a standardized procedure for administering, scoring and interpreting
- There is typically not much flexibility with regard to test administration, though accommodation for test taking barriers can be made
- For example, on the WAIS-IV, which is made up of 10 subtests, the examiner can substitute the Block Design subtest (which requires use of hands/motor movement) if the person has restricted motor ability, for a subtest that measures the same construct (in this case, fluid intelligence)

Test scoring

- Tests yield raw scores (raw scores would not typically be helpful in clinical assessment, though there are exceptions)
- These are added either by the examiner counting the responses/items or entering the examinees responses into a computer and letting the computer do the work
- Some of the work of scoring tests is difficult and tedious, some is very straightforward
- For, example, in scoring the Vocabulary subtest of the Wechsler Adult Intelligence Scale, the examinees responses has to be assigned either a 0, 1 or 2!
- In scoring the MMPI by computer, the examiner simple enters the person's age and gender, then types whether the responses were True or False

Test interpretation

After test scoring, the next step is to interpret the scores
The first step, is to determine if the test is a valid administration
The examinees effort or level of cooperation may play a role
There may have been less than optimal testing conditions (for example testing in a jail setting, where there was intermittent noise levels)
Personality tests usually have validity scales; these help determine whether the results can be trusted or whether there should be some caution used in the interpretation

Test Reporting and Assessment

- ◆ Assessment and testing are two different activities
- ◆ An assessment usually includes the following:
 - ◆ Interviewing
 - ◆ Testing
 - ◆ Collateral information
 - ◆ Review of professional information
 - ◆ Putting together a report
 - ◆ Offering a diagnosis based on data
 - ◆ Offering a treatment plan based on data, diagnosis and available resources

Cultural Issues

- Cultural and Language issues are important at all stages of testing
- For example, a person who has only recently arrived to the US (no exposure to US culture), would not be familiar enough with US customs, may be language barriers, would not fit with norms
- Would not be appropriate to give the MMPI, an IQ test
- There are Spanish versions of the MMPI, IQ tests, but these should be administered by a psychologist fluent in the language

Hypothesis

- Psychologists should start with the "null hypothesis", that is, there is nothing at wrong with the client
- This is usually not too easy to do, since there has been a referral made to the psychologist, to do an assessment
- The psychologist has to strive to build a hypothesis, test it out, reject it when new information comes in, engage in contrafactual and critical thinking and guard against confirmation bias

Bad use of testing

Examiner does not administer test correctly
Does not score test correctly
Does not read the test manual
Misuses test scores and inflates scores ability to predict
Makes recommendations based on a single score or single test
Does not use test for the test's intended purpose

Assessment reports

- Test reports should be transparent
- Reports need to reflect the effort of the examiner to address the referral question
- The "art" of the assessment process is integrating the data so the results are communicated in a way your grandmother should be able to understand
- This assumes your grandmother is not a professional psychologist, of course, but...she could be...and in that case, if she is.... she should still be able to understand it!

Personality

- Personality...



Self report instruments

- Self report instruments are those questionnaires or inventories that are given to a client to complete in the office
- These cannot be taken home
- These are untimed
- There are broadband instruments which tend to cover multiple problem areas (MMPI, PAI) and instruments that only cover a certain topic, for example, trauma, parenting, depression, anxiety, anger, substance abuse
- Optimally, each evaluation contains one broadband instrument, but sometimes, given the client's function (low reading ability, disorganization), brief questionnaires are the way to go

MMPI-2 (Minnesota Multiphasic Personality Inventory)

- First published in 1942 by Hathaway and McKinley
- Designed for routine diagnostic assessments
- Empirical keying approach
- 724 Minnesota normals and 221 psychiatric patients
- Normative sample was Caucasian, 8th grade education, a convenience sample
- Originally 8 clinical scales plus validity scales
- Many more scales have been added over the years

MMPI-2

- Empirical or criterion keying approach to test development
 - Identify a criterion group (e.g., people diagnosed with schizophrenia)
 - Identify a comparison group (e.g., persons with no mental illness)
 - Administer many, many test items to both groups
 - Identify a group of items that discriminates the two groups, i.e., items endorsed more frequently by the criterion group
 - This group of items becomes the schizophrenia scale

MMPI-2

- ◆ 1980s began restandardization
- ◆ Items changed to reflect change in culture, trends, elimination of outdated words, terms
- ◆ The new sample did not exclude people undergoing medical treatment, like originals
- The MMPI-2 normative sample consists of 2,600 individuals (1,138 men; 1,462 women), age 18 or older
- There is a MMPI-A, for youth 14- to 18 years of age

MMPI-2

- The number of items on the MMPI-2 is 567
- All items are marked True or False
- The reading level varies from about the 6th to the 9th grade
- Items can be read by the client, but not to the client
- The MMPI can be administered by audio, so the client can listen to a CD player or by using a laptop
- The exam typically takes 90 minutes or longer

MMPI-2

- Alice was a 40 year old female referred for a assessment with a rule out diagnosis of an anxiety disorder
- It was clear she had anxiety during the interview; she also appeared to be many obsessive features
- She was given the option of the audio (which takes two hours) or the written administration
- She took the written version and completed it in 4 hours
- A reading disability had been ruled out
- She had many conflicts about what answer to give and spent an excessive amount of time ("I'm stuck on this one again" she would say) with each answer
- Her approach mirrored her everyday struggles with making decisions and helped inform her treatment plan

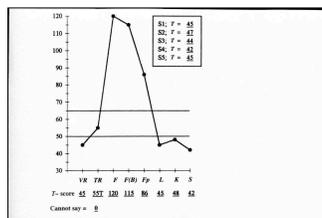
MMPI-2

- Once the items are completed by a client, the MMPI is scored
- Most psychologists score the MMPI by computer
- There are two scoring services
- Caldwell Report
- Pearson
- There is the option to score with templates
- Computer scoring is recommended to reduce potential examiner error and for efficiency purposes

Validity scales

- There are several validity scores; most contemporary tests have learned from the MMPI that the validity scales are very helpful in determining the usefulness of the results, personality style and treatment readiness
- Measured are consistency of response style, level of openness to discussing problems, type of defensiveness or minimization (such as denial, naive denial, need to present as adequate), tendency to exaggerate symptoms
- The MMPI can help determine treatment readiness, for example, if a person is presenting themselves as virtuous, very adequately controlled, out of touch with their feelings, or extremely overwhelmed...these factors can be taken into consideration when formulating a treatment plan

Exaggerated Profile



Scales

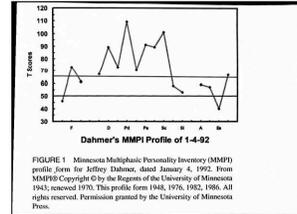
- There are 10 clinical scales with scales measuring various clinical syndromes ranging from psychogenic pain to mania to shyness
- The scales themselves do not always represent what they are named
- For example, scale 4 was named the psychopathic deviate scale; this scale can measure psychopathy (a whole other training!) but two people can score high on scale 4, and the scale can have two different meanings
- The MMPI also has many other scales, which can be helpful, relative to treatment planning
- For example, there are measures of anger, family problems, substance abuse risk, work related problems and measures of coping style, such as internalizing versus externalizing



The case of Jeffrey Dahmer

- Produced two valid MMPI-2 profiles
- He was tested just before his sanity trial in 1992
- His extremely high 4 scale was noteworthy, as well as his very high 8 scale (scale 8 was named Schizophrenia)
- These two scales in combination are not uncommon in psychiatric populations, but it would be very rare to have scales as high as Dahmer's

Dahmer MMPI



Interpretation of Dahmer profile

- "Standard interpretive sources have described similar patients as odd and peculiar in thinking and behavior and as distant and mistrustful in relations with others despite strong needs for attention and affection. They see others as hostile, rejecting, and unreliable. Sexual deviation, the confusion of sexuality with aggression, defective empathy, and difficulties in the expression and control of anger are chronic problems. Crimes committed by such patients tend to be poorly planned and executed and may involve bizarre and violent behavior." Nichols (2006)

The case of Mr. Monk



- If a client obtains two or more scales that cluster together, these are referred to as code types
- Detective Adrian Monk would probably be a classic 2-7 code type on the MMPI, a very common profile in clinical settings
- Mr. Monk has excessive tension, obsessions and worry
- Scale 2 indicates his ongoing depression, but even higher would probably be his scale 7 (tension, worry and obsessive thinking)

Treatment plans

- Dahmer did obtain a valid MMPI and possibly Mr. Monk would produce a valid MMPI profile
- While Dahmer would not be treatable, Mr. Monk is...and is in therapy (with slow progress, using therapy as reassurance with some growth)
- That's because Mr. Monk has anxiety and depression as indicated by his MMPI, common psychotherapy presentations (though his OCD is more rare)

Treatment plans: Substance Abuse



The MMPI has three scales that help assess substance abuse. These scales look at both direct endorsement of substance abuse items and indirect factors such as risk taking, impulsive behavior. The PAI has a Drug Scale and an Alcohol Scale. Higher scores suggest dependence versus misuse/abuse. There is also a way to determine under reporting of both Drug and Alcohol problems.

Personality Assessment Inventory (PAI)

- PAI is a self-report instrument.
- 344 items
- Reading level of items is about fourth grade
- Administered by paper pencil or audio
- 4 validity scales.
- 11 clinical scales.
- 5 treatment scales related treatment and case management.
- 2 interpersonal scales.
- Has ranked fourth in personality measures used by clinicians

PAI

- Developed between 1987 and 1991
- development used rational and quantitative method of scale development.
- Constructs assessed were selected on the basis of the stability of their importance within the concepts of mental disorder and significance in clinical practice.

PAI

- During item generation, for each syndrome, the literature on each construct was examined identify the components that were most central to the definition of the concept; the items were written with the goal of providing and assessment of these components
- The MMPI and PAI differ because of this rational emphasis purely empirical development
- The PAI scales are designed to measure the constructs that are represented by the names of the scales

PAI

- The PAI seeks to sample information, which is relevant to important clinical constructs such as psychosomatic, antisocial or borderline personality problems
- It attempts to sample the breadth and the depths or intensity of each construct
- Depth is assessed by using a Likert scale versus a true or false format.
- Responses to items range from very true, slightly true, mainly true and false

PAI norms

- PAI does not use separate norm for men and women.
- Items were selected to have the same meaning regardless of gender.
- Men tend to score, on average, five points higher on the antisocial scale in the alcohol scale

PAI Scales Summary

Inconsistent Responding	Borderline Personality
Negative Impression Management	Antisocial
Positive Impression Management	Alcohol and Drug Scales
Somatic Complaints	Aggression
Anxiety	Suicidal Ideation
Anxiety Related Disorders	Stress and Social Support
Depression	Treatment Rejection
Mania	Warmth and Dominance
Paranoia	
Schizophrenia	

Specific Problem Area Assessment

- The MMPI and PAI have validity measures; some tests do not have great validity measures and tend to have high face validity (it is very easy to recognize your reporting problems)
- Cognitive Distortion Scales- a 40 item measure can be helpful in understand mental health issues but also helps understand victims of violence
- NAVACO Anger and Provocation Scale- is a measure of just what it says (ages 9 years old to 84 years old) and help understand anger regulation, justification and coping strategies

Specific Problem Area Assessment

- Joan was a 28 year old female referred for a parenting evaluation
- She has a trauma background; she was too restless, anxious and disorganized to concentrate on a broadband test
- She took the CDS; this showed an elevation on a scale measuring Preoccupation with Danger and Helplessness
- This helped with further understanding how her remote stress was functioning as a current stress and how she felt ineffective (high helplessness score) but likely not suicidal (low Hopelessness score)

Parenting Assessment

- Observation of child/parent interaction
- Assessment of parenting knowledge
- Adult Adolescent Parenting Inventory-2
- Risk assessment for child physical abuse
- Child Abuse Potential Inventory

Adult Adolescent Parenting Inventory-2

- Norms for teen parents, parents without any parenting training
- Male and female norms
- 40 items, using a rating scale
- Uses Stanine scores (1-10)
- Inappropriate expectations, empathy, alternatives to corporal punishment, role reversal, power and independence
- Low scores = deficit
- High scores = better knowledge

AAPI-2

- Can help determine if parenting problems are present, cannot entirely rule them out
- Should take in consideration of culture in evaluating responses

Parent With Knowledge Deficits

- William was a parent of four young children
- CPS found home to be in disarray, children not going to school, William not supervising children, delegated parenting to 9 year old daughter
- William obtained the following scores on the AAPI-2:
 - Inappropriate expectations 2
 - Empathy 5
 - Alternatives 5
 - Family roles 1
 - Power 5

Child Abuse Potential Inventory

- CAPI is a 160 item inventory
- Specifically identified for parents at risk for child physical abuse in CPS populations
- Use with parents 18 and up
- Parents respond with agree or disagree
- Has good construct validity (measures what it is intended to measure)
- Has abuse scale: when elevated, few false positives
- Parents who respond to treatment, abuse score lowers

CAPI

- Larry was a 40-year-old parent of two young children, and had been referred by a local agency, due to his history of spousal abuse, which was reported to be severe, and had taken place against one of kid's mothers
- There were concerns that he was callous and abrasive with his children during parenting time visits
- Larry completed a valid CAPI; he obtained an elevated Abuse Scale. The scale suggests he exhibited a history of child abuse himself (which was confirmed by his psychosocial history), showed features of being touchy, over-reactive and irritable, with lower frustration tolerance

Cognitive and Neuropsychological

- IQ tests use standard scores
- The mean is 100
- The Wechsler Scales are the most commonly administered
- There are three current versions
- Wechsler Adult Intelligence Scale-IV (WAIS-IV)
- Wechsler intelligence scale For Children-IV him
- WPPSI-III

Things to note about IQ tests

- Wechsler Scales take anywhere from 45 minutes to 90 minutes to administer
- Results can help with treatment planning
- High IQ = solve problems quickly, generally have better memory and learning potential, can be a skilled worker
- Low IQ solve problems slower, have less developed memory and learning potential, typically are unskilled workers

Other things to note about IQ

- IQ tests, just like other tests, have error built into them
- With the Wechsler scales, psychologists should report more than just the score; they need to report the "margin of error" or confidence interval
- This is a +/- band or range
- This helps the reader understand that the Full Scale IQ is not score that should be idolized

Brief IQ tests

- There are brief IQ tests that work well in assessing parents, medically fragile parents, people in jail
- Angie was a single parent, referred because of her overt hostility and inclination to be combative with treatment providers as well as her neglect and poor supervision of her kids
- She called the office 7 times for directions
- Her Composite score on the Shipley-2 was 58, which is below the first percentile
- 99.99 percent of Angie's peers scored better
- Or, she scored better than .01 percent of her peers
- Either way, her cognitive abilities are quite compromised and offers an explanation of limitations, deficits in judgment and insight

Neuropsychological Screens

- Clients with a history of head injury or CNS insults should probably have a referral for a neuropsychological screening
- Such screening can be helpful in evaluating:
 - Attention
 - Language
 - Memory
 - Pre-frontal or Executive functioning



Neuropsychological Screens

- One of the best tools a psychologist has, are interviewing skills
- A thorough interview and history can be useful, but you have to asked the right questions
- A review of a client symptoms, family history, history of head injury or head injuries, domestic violence history (included here would be child abuse history) -such injuries may be underreported, medical history, educational history
- Information from relatives can be a good source of collateral information

Neuropsychology

- Screening measures administered by psychologists are different than longer, neuropsychological tests which are given by neuropsychologists and are often very lengthy, expensive and not always easily to access for clients
- A mental status exam can be very helpful in determining issues such as difficulties with orientation, attention, reasoning, visual motor skills, judgment and planning
- A mental status exam can be scored, and uses a cutoff score, to help identify risk for dementia or neuropsychological compromise

Neuropsychological screening tests

- Wide Range Assessment of Memory and Learning
- Comprehensive Trail Making Test
- Stroop Test
- Controlled Oral Word Association Test
- Perceptual Reasoning subtests from the Wechsler scales
- Working Memory subtests from the Wechsler scales

When to refer for a psychological assessment

- For risk assessment
- To check how serious a mental health problem is, with an objective measure
- To view the problem from an unbiased source/method
- To assess the client's capacity for insight, for example to rule out a problem with IQ or severe personality disorder
- To see how mental health problems may interfere with parenting
- To better understand a client's unpredictable or confusing responses to treatment

What to include in your referral information

- Your observations of the client
- Any history of response to services
- Collateral information from other assessments, for example substance abuse assessments, psychiatric hospitalizations
- Specific information you hope to learn from the assessment (referral question)
- You can always call for assistance with developing a referral question.

Summary

- In summary, psychological testing can be helpful to understand a client's psychological functioning, treatment needs and readiness for treatment
- Not every psychologist practices assessment psychology, so use the information from this presentation to ask appropriate questions of the psychologist
- Psychological tests are imperfect measures, but better than tea leaves
- You can always call or email me with any questions: 269-381-4552 or robert.griffith@wmich.edu

References

- Butcher, J.N., (2005). A beginner's guide to the Minnesota Multiphasic Personality 2nd Edition. Washington DC: American Psychological Association.
- Butcher, J.N., Graham, J.R., Ben-Porath, Y.S. Tellegan, Y.S. Dahlstrom, W.G., & Kaemmer, B. (2001). Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring and interpretation (Rev. Ed.). Minneapolis: University of Minnesota Press.
- Groth-Mamat, G. (2009). Handbook of psychological assessment. 5th Edition. Hoboken, N.J.: John Wiley & Sons, Inc.
- Nichols, D (2006). Tell me a story: MMPI responses and personal biography. In the case of a serial killer. Journal of Personality Assessment, 86 (3) 242-262.
- Salkind, N.J. (2006). Tests & measurement for people who (think they) hate tests & measurement. Thousand Oaks, CA: Sage Publications.
- Wright, A.J. (2011). Conducting psychological assessments: a guide for practitioners. Hoboken, N.J.: John Wiley & Sons, Inc.