

Ch. 7: Test Administration

- Theory - what affects test scores?
- The Examiner and the Subject
 - Relationships between Examiner and Subject
 - Race, Language of subject
 - Examiner Training
 - Expectancy effects / Reinforcement
 - Computer-administered testing
 - Subject Variables
- Behavioral Assessment Methodology
 - Reactivity, Drift, Expectancy
- Deception & Detection of Malingering

473

Psychology 402 - Fall 2014 - Dr. Michael Diehr

474

Psychology 402 - Fall 2014 - Dr. Michael Diehr

What is the “protocol”?

- Simple view :
 - A test is just the collection of test items and grading rules and norms (What)
- Must consider the entire setting:
 - Set: Why?
 - Setting: Where/When
 - Examiner: Who?
 - Method of administration: How?
- With tests, often the “What” is specified carefully but the others are left open to interpretation...

475

Psychology 402 - Fall 2014 - Dr. Michael Diehr

476

Psychology 402 - Fall 2014 - Dr. Michael Diehr

What affects test scores?

- Simple View / Classical Test theory
 - $X = T + E$
 - Observed Score = True Score + Random Error
- Modern View
 - Error - is not always random
 - Error - comes from both subject AND protocol

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Rapport & IQ scores

- Feldman & Sullivan (1960) : children taking WISC
 - Neutral condition
 - High Rapport condition
 - Result: IQ scores for 5th thru 9th graders went up (122 vs 109)
- Review by Fuchs & Fuchs (1986) of 22 studies across 1500 students: Performance increased 0.28 SD with familiar examiners (up to 0.5 SD in lower SES students)
- Question: given these results, is cross-cultural testing fair?

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Race of Tester & Subject

- Common belief that IQ testing administered to minorities, by examiner who is not minority, is unfair
- Satler (2002,2004) reviews studies, finds evidence is minimal (only 4 of 29 studies), concludes it is a “myth”
- What is basis for this belief?
 - Ethnic minorities do tend to score lower on standardized testing, and tests are generally administered by non-minority testers
- Possible instance of correlation ≠ causation?

477

Psychology 402 - Fall 2014 - Dr. Michael Diehr

478

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Race of Tester & Subject 2

- Some studies do show effects
- Effects tend to be larger when
 - testing protocol is more flexible
 - testers are less-well trained
- Explanation?

Language issues

- Can not assume that language is not an issue
- Translating a test into another language may change reliability & validity
- More on this topic later

479

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Training of Testers

- Giving a test properly can be difficult
- Patterson et al. (1995) showed numerous errors in administering the WAIS-R, performance only improving after 10 administrations. (Graduate students often get only 4 administrations)
- Behavior of Testers is largely unspecified.
- Example : Do you say “yes” or “good job” when an answer is correct? Do you say “keep trying” if a person appears to be giving up easily?

480

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Expectancy Effects

- Definition: tendency for a person to find evidence that supports a pre-existing hypothesis
- Theory:
 - (A) selection bias in collecting data -- ignore data that seems wrong, accept data that fits your theory
 - (B) actually changing the environment -- encourage desired behavior by subtle or overt prompting

481

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Expectancy Effects: Rosenthal

- Rosenthal (1966)
 - Give students a set of faces to rate
 - Rate pictures of faces on “Success” or “Failure”
 - Half of students were led to believe the faces were of Successful people. Both groups received same pictures.
 - Result: ratings from the experimental group were about 1 point higher (on a 20 point scale)
 - Conclusion: expectation influences judgement
 - Effects seen when rating non-humans (e.g. rats)

482

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Expectancy Effects: Testing

- Sattler et al. (1970) showed expectancy effects when rating an ambiguous response on an IQ test.
 - Give ambiguous response to various raters
 - Tell half the raters it is a “smart” child.
 - Results: “smart” children scored better.
- Sattler (1998) showed this result even when the test answer was not ambiguous.
- Review of literature shows inconsistency. Results on the whole tend to be small.
- Conclusion: real phenomenon, design tests to avoid by specific scoring rules

483

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Lawyers Guns and Money

- Reinforcements shape behavior. Can reinforcement change test results? IQ tests?
- Theoretical issue: if money, candy or praise can improve IQ score, what exactly is IQ a measurement of?
- Research is complex and inconsistent : Bergan et al. (1971) -- boys responded to tokens, response to praise was mixed (girls: performance improved but slowed; boys went faster.)
- “Nice Job little brother” : Terrel et al. (1978) showed big results when providing “culturally relevant feedback”

484

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Are your intestines too long?

- As expected, effects tend to be higher on tests without a right or wrong answer, or for subjects that may be uncomfortable
- Cannell (1974) : “yes” answers to physical symptoms increased when interviewer gave approving nod.
- Yes answers increased to nonsense questions: “Do the ends of your hair itch” and “Are your intestines too long?”

485

Psychology 402 - Fall 2014 - Dr. Michael Diehr

The Humans are Dead

- Human interviewers can change the performance of subjects taking tests
- Question: use robots instead?
- Pros:
 - complete standardization
 - adaptive testing
 - precision of timing
 - cost effective
 - patience
 - bias reduced
 - encourage socially undesirable responses?

489

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Robots make better testers?

- Resenfeld et al. (1989) found subjects preferred computer-test to paper-and-pencil test
- Lock & Gilbert (1995) gave subjects the MMPI either via computer, paper & pencil, or interview, and found people to be most revealing of undesirable information with the computer. Also, people rated the computer as most pleasant.
- Studies show computers at least as reliable as humans
- Issues about validity:
 - administration vs. scoring vs interpretation.
 - Testing vs. Assessment

490

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Automated Testing Issues

- Boundaries of competence? “Psychologists provide services...only within the boundaries of their competence...”
- Scientific Basis? “Psychologist’s work is based on established scientific [...] knowledge of the discipline”
- Delegation of Work: “Psychologists who delegate work...authorize only those responsibilities that such persons can be expected to perform competently...”
- Use of Assessments: “Psychologists use assessment instruments whose validity and reliability have been established for use...”

492

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Automated Testing Issues 2

- Use of Assessments: “Psychologists use assessment instruments whose validity and reliability have been established for use with members of the populations tested.”
- Assessment by Unqualified Persons: Psychologists do not promote the use of psychological assessment techniques by unqualified persons, except when such use is conducted for training purposes with appropriate supervision.

493

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Automated Testing Issues 3

- “When interpreting assessment results, including automated interpretations, psychologists take into account the purpose of the assessment as well as the various test factors, test-taking abilities, and other characteristics of the person being assessed, such as situational, personal, linguistic, and cultural differences, that might affect psychologists’ judgments or reduce the accuracy of their interpretations

494

Psychology 402 - Fall 2014 - Dr. Michael Diehr

“Subject Variables”

- Motivation
- Anxiety
- Illness
- Medications
- Hormones
- Sleep
- etc...

497

Psychology 402 - Fall 2014 - Dr. Michael Dohr

Army Alpha and Beta

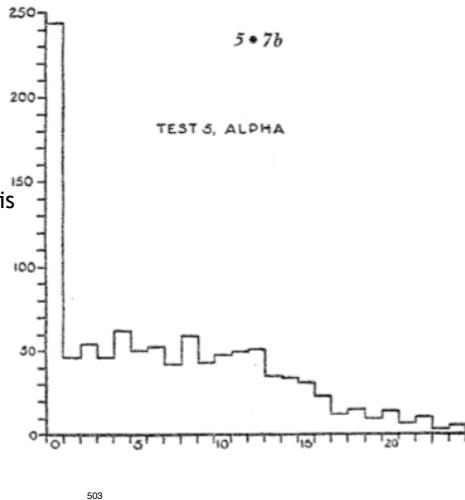
- Sample questions
- Sample administration protocol
- Results

498

Psychology 402 - Fall 2014 - Dr. Michael Dohr

Army Alpha Results

- What is a test like this measuring?



503

Psychology 402 - Fall 2014 - Dr. Michael Dohr

Review

- Theory - what affects test scores?
- The Examiner and the Subject
 - Relationships between Examiner and Subject
 - Race, Language of subject
 - Examiner Training
 - Expectancy effects / Reinforcement
 - Computer-administered testing
 - Subject Variables

506

Psychology 402 - Fall 2014 - Dr. Michael Dohr

Behavioral Assessment

- aka “Functional Assessment”
- Work samples, on-the-job testing, “in situ” / “in vivo”
- More active role of psychologist / observer / rater can lead to bigger problems with accurate measurement
- Reactivity
- Drift
- Expectancies

507

Psychology 402 - Fall 2014 - Dr. Michael Dohr

Reactivity

- Reliability of observers is highest when the observers are being observed
- Reid (1970) : observer accuracy dropped 25% when told their work would not be measured
- Methods: random sampling, covert sampling
- Measures of test Reliability are often done in ideal situation, not everyday situation.

508

Psychology 402 - Fall 2014 - Dr. Michael Dohr

Drift

- Observers can be trained to certain level of accuracy, but their performance tends to change slowly over time.
- A 9/10 rating when you first started may only be an 8/10 now.
- Drift can happen on individual or group basis.
- Group drift especially hard to counteract, since the group members tend to support each other's ratings.
- Drift is frequently ignored: 17% of studies even report the # of raters. 10% documented the training of raters. 5% tested for drift.

509

Psychology 402 - Fall 2014 - Dr. Michael Diehr

510

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Expectancies in Beh. Obs.

- Expectancy effects, in Behavioral Observation situations, are similar to those seen in Testing
- Effects are subtle, small, but real and can be a significant problem in some contexts
- Effects seem to be largest when Observer is rewarded for reporting certain behaviors.

Deception

- Most people, even trained professionals, are remarkably poor at detecting lying/deception (possible exception - Secret Service agents?)
- Polygraph "Lie Detector" tests are neither reliable nor valid
 - Example: correlation between honesty test and thefts : $r = 0.13$, ($r^2 = .02$) meaning about 2% of variance is explained.
 - Over 95% false positive rate
- Some belief that participation by Psychologists in such testing is violation of ethical principles (Camara & Schneider 1994)
- Only a few countries use them at all (e.g. not used in Europe)

512

Psychology 402 - Fall 2014 - Dr. Michael Diehr

U.S. DEPARTMENT OF LABOR

EMPLOYMENT STANDARDS ADMINISTRATION

Wage and Hour Division
Washington, D.C. 20210



NOTICE

EMPLOYEE POLYGRAPH PROTECTION ACT

The Employee Polygraph Protection Act prohibits most private employers from using lie detector tests either for pre-employment screening or during the course of employment.

PROHIBITIONS

Employers are generally prohibited from requiring or requesting any employee or job applicant to take a lie detector test, and from discharging, disciplining, or discriminating against an employee or prospective employee for refusing to take a test or for exercising other rights under the Act.

EXEMPTIONS*

Federal, State and local governments are not affected by the law. Also, the law does not apply to tests given by the Federal Government to certain private individuals engaged in national security-related activities.

Psychology 402 - Fall 2014 - Dr. Michael Diehr

The Act permits *polygraph* (a kind of lie detector) tests to be administered in the private sector, subject to restrictions, to certain prospective employees of security service firms (armored car, alarm, and guard), and of pharmaceutical manufacturers, distributors and dispensers.

The Act also permits polygraph testing, subject to restrictions, of certain employees of private firms who are reasonably suspected of involvement in a workplace incident (theft, embezzlement, etc.) that resulted in economic loss to the employer.

EXAMINEE RIGHTS

Where polygraph tests are permitted, they are subject to numerous strict standards concerning the conduct and length of the test. Examinees have a number of specific rights, including the right to a written notice before testing, the right to refuse or discontinue a test, and the right not to have test results disclosed to unauthorized persons.

ENFORCEMENT

The Secretary of Labor may bring court actions to restrain violations and assess civil penalties up to \$10,000 against violators. Employees or job applicants may also bring their own court actions.

ADDITIONAL INFORMATION

Additional information may be obtained, and complaints of violations may be filed, at local offices of the Wage and Hour Division. To locate your nearest Wage-Hour office, telephone our toll-free information and help line at 1-866-4USWAGE (1-866-487-9243). A customer service representative is available to assist you with referral information from 8am to 5 pm in your time zone; or if you have access to the internet, you may log onto our Home page at www.wagehour.dol.gov.

THE LAW REQUIRES EMPLOYERS TO DISPLAY THIS POSTER WHERE EMPLOYEES AND JOB APPLICANTS CAN READILY SEE IT.

*The law does not preempt any provision of any State or local law or any collective bargaining agreement which is more restrictive with respect to lie detector tests.

U.S. DEPARTMENT OF LABOR
EMPLOYMENT STANDARDS ADMINISTRATION
Wage and Hour Division
Washington, D.C. 20210

WH Publication 1462
June 2003

514

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Lie Detector Tests

- Prohibited by employee Polygraph Protection Act of 1988 (EPPA).
- "Employers generally may not require or request any employee or job applicant to take a lie detector test, or discharge, discipline, or discriminate against an employee or job applicant for refusing to take a test or for exercising other rights under the Act."
- Exceptions -- security firms and pharmaceutical manufacturers, and government.
- Not admissible in court of law (Frye Rule from 1923)

515

The low base rate / False Positive Problem

516

Decision Making & Errors

		The Real World	
		Guilty	Innocent
You Decide	Guilty	True Positive 1- β Power	False Positive Type I Error α Alpha
	Innocent	False Negative Type II Error β	True Negative 1- α

Psychology 402 - Fall 2014 - Dr. Michael Diehr

517

Psychology 393 - Cognitive Lab - Spring 2014

The low base rate / False Positive Problem

- Scenario
 - 10,000 people tested
 - 10 are actually spies
 - Lie Detector Test
 - 84% accuracy (theoretical)
 - 16% false positive rate

518

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Example

		The Real World	
		Guilty	Innocent
You Decide	Guilty	True Positive 8	False Positive 1598
	Innocent	False Negative 2	True Negative 8392

Psychology 393 - Cognitive Lab - Spring 2014

519

“Get a good night’s sleep”

- NAS concluded that if 10,000 employees (of whom 10 were spies) were given a polygraph:
 - 8 spies would fail the test
 - 1598 non-spies would fail the test
 - Roughly 99.6% of those failing the test would be False Positives
 - This assumes a very optimistic 84% accuracy (actual accuracy is probably much worse)
- Notorious spies not being caught:
 - Aldrich Ames passed two polygraphs.
 - His advice? “Get a good night’s sleep...”

520

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Lie Detector Tests 2

- If they are not valid & reliable, and prohibited by law, what good are they?
- Common belief that they are accurate means they can serve as a punishment, and an inducement to confession.
- In other words, the fear of being caught causes such severe anxiety that a person may choose to confess (even sometimes, to a crime not committed).
- Test can be fairly easily beat with simple training.
- What sorts of people are likely to lack knowledge and be susceptible to these tests?

521

Psychology 402 - Fall 2014 - Dr. Michael Diehr

Detection of Malingering

- There are sometimes benefits of performing poorly on a test (disability, forensic, military, etc.)
- Often called “faking bad”
- On certain esoteric tests, it’s very hard for an untrained person to know what a “normal” performance level is.
- Malingering tests give false feedback, which can encourage a person faking bad to perform worse than people with actual injury. In some cases, perform worse than chance.
- Note: not all such performance is intentional. Possible for patient to believe in their illness.

523

Hiscock Forced-Choice Procedure

524

Psychology 402 - Fall 2014 - Dr. Michael Dishr