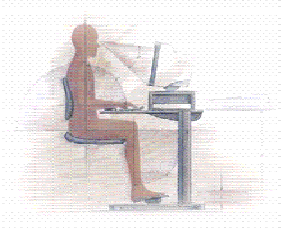
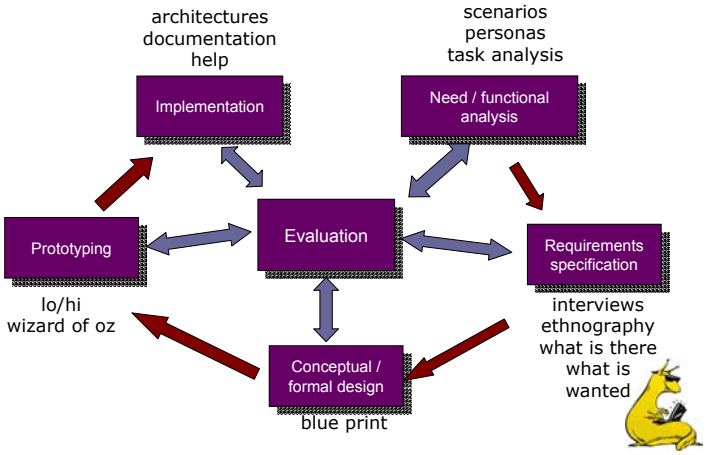
CMPE 233: Human Factors



Evaluation

User-centered Design



Evaluation

- Concerned with gathering data about the usability/ usefulness of
  - a design or product
  - by a specific group of users with certain limitations/abilities
  - for a particular activity/task
  - in a specified environment or context
- Formative – early, meshed closely with design, guides the design process
  - Predict usability of product or aspect of product
  - Check design team's understanding of user requirements
  - Test out ideas quickly and informally
- Summative – late, judgment about the almost finished product
  - identify user difficulties / fine tune
  - improve an upgrade of product

Reasons for doing evaluations

- Understanding the real world
  - How to employ in workplace?
  - Better fit with work environment?
- Comparing designs
  - compare with competitors or among design options
- Engineering towards a target
  - x% of novice users should be able to print correctly on first try
- Checking conformance to a standard
  - screen legibility, accessibility.

Case Study: Air Traffic Control

- CAA (Civil Aviation Authority) in the UK, 1991
- Original system -- data in variety of formats
  - analog and digital dials
  - CCTV, paper, books
  - some line of sight, others on desks or ceiling mountings outside view displays
- Goal: integrated display system, as much info as practical on common
- Major concern: safety
- Evaluate controller's task
  - want key info sources on one workstation (winds peed, direction, time, runway use, visual range, meteorological data, maps, special procedures)

Air Traffic Control, continued

- Develop first-cut design (London City, then Heathrow)
- Establish user-systems design group
- Concept testing / user feedback
  - modify info requirements
  - different layouts for different controllers and tasks
  - greater use of color for exceptional situations and different lighting conditions
  - ability to make own pages for specific local conditions
  - simple editing facilities for rapid updates
- Produce upgraded prototype, "Road Show" to five airports, develop system specification
- Build and install system: Heathrow in 1989, → get further needs; others 1991

Case Study: Forte Travelodge

- System goal: more efficient central room booking
- IBM Usability Evaluation Centre, London
- Evaluation goals:
  - identify and eliminate problems before going live
  - avoid business difficulties during implementation
  - ensure system easy to use by inexperienced staff
  - develop improved training material and documentation
- Setup a usability lab to study users
  - Similar to TV studio: mic, audio, video, one-way mirror
  - set up to resemble Travelodge reception area, attempt to be non-threatening

Travelodge: Procedure

- Developed set of 15 common scenarios, enacted by cross-section of staff over 8 half-day sessions
- Emphasize that evaluation is of system not staff
- Debriefing sessions after each testing period, get info about problems and feelings about system and doc
- Areas of interest:
  - system navigation, speed of use
  - screen design: ease of use, clarity, efficiency
  - effectiveness of onscreen help and error messages
  - complexity of keyboard for computer novices
  - effectiveness of training program
  - clarity and ease-of-use of documentation
- New system: higher productivity, low turnover, faster booking, greater customer satisfaction

Evaluation Methods

- Observing and monitoring behavior
  - field or lab, observer takes notes / video
  - keystroke logging / interaction logging
- Collecting users' opinions
  - interviews / surveys
- Experiments and benchmarking
  - semi-scientific approach (can't control all variables)
- Interpretive Evaluation
  - informal, try not to disturb user; user participation common
  - includes participatory evaluation, contextual evaluation
- Predictive Evaluation
  - predict problems users will encounter without actually testing the system with the users
  - keystroke analysis or expert review based on specification, mock-up, low-level prototype

Why Use Different Methods?

- Information requirements differ
  - pre-design, iterative design, post-design, generalizable knowledge…
- Information produced differs
  - outputs should match the particular problem/needs
- Cost/benefit of using a certain method
- One method's strength can complement another's weakness
  - no one method can address all situations
- Constraints
  - may force you to chose quick and dirty discount usability methods

How Can We Compare Methods?

- Type of information (qualitative vs. quantitative)
- Relevance
  - does the method provide information to our question / problem?
- Setting
  - is it important that the system be evaluated in-context?
- Generalization
  - how well can I generalize the information produced to other situations?
- Repeatability
  - would the same results be achieved if the test were repeated?

How Can We Compare Methods?

- Quickness
  - can I do a good job with this method within my time constraints?
- Cost
  - is the cost of using this method reasonable for my question?
- Equipment
  - What special equipment / resources required?
- Personnel, training and expertise
  - What people / expertise are required to run this method?
- Validity
  - External validity: can the results be applied to other situations?
  - Internal validity: do we have confidence in our explanation?

How Can We Compare Methods?

- Subject selection
  - how many do I need, who are they, and can I get them?
- Scope of subjects
  - is it good for analyzing individuals? small groups? organizations?
- Control
  - do I need to control for certain factors to see what effects they have?
- Cross-sectional or longitudinal
  - is it important that changes over time are measured?
- Support
  - are there tools for supporting the method and analyzing the data?
- Comparative
  - can I use it to compare different things?

How Can We Compare Methods?

- Does the test measure something of relevance to usability of real products in real use outside of lab?
  - Some typical reliability problems of testing vs. real use
    - non-typical users tested
    - tasks are not typical tasks
    - physical environment different
    - social influences different

Direct Observation

- Difficulties:
  - people "see what they want to see"
  - "Hawthorne effect" -- users aware that performance is monitored, altering behavior and performance levels
  - single pass / record of observation usually incomplete
- Useful: early, looking for informal feedback, want to know the kinds of things that users do, what they like, what they don't
- Know exactly what you're looking for → checklist
- Want permanent record: video, audio, or interaction logging
- Ethnography: Immerse in situation you want to learn about (participant observer, privileged observer)

Indirect Observation

- Alleviates some difficulties of direct observation
- Can be synchronized with keystroke logging or interaction logging
- Problems:
  - effort required to synchronize multiple data sources
  - time required to analyze
  - users aware they're being filmed → set up and leave for several days, they get used to it
- Virtual ethnography (netnography)
  - Like ethnography but for technologically mediated interactions in online networks and communities
  - Used to observe blogs, web-rings, chat, SMS, game communities, bulletin boards, and mailing lists

Analyzing video data

- Task-based analysis
  - determine how users tackled tasks, where major difficulties lie, what can be done
- Performance-based analysis
  - obtain clearly defined performance measures from the data collected (frequency of task completion, task timing, use of commands, frequency of errors, time for cognitive tasks)
  - classification of errors
  - repeatability of study
  - time (5:1) -- tools can help
- Slight variation → video of screen interaction
  - complement other data collection
  - important to cross check user comments

Verbal protocols

- User's spoken observations, provides info on:
  - what user planned to do and their mental model
  - user's identification of the terms they use to refer to objects or actions
  - reactions when things go wrong, tone of voice, subjective feelings about activity
- "Think aloud protocol"
  - user says out loud what s/he is thinking while working on a task or problem-solving
- Post-event protocol (retrospective testing)
  - users view videos of their actions and provide commentary on what they were trying to do
  - less intrusive, important for time- and error-sensitive tasks or for people with ADHD

Interviews

- Structured interviews
  - predetermined questions, asked in a set way
  - no exploration of individual attitudes
  - structure useful in comparing responses, claiming statistics
- Flexible interviews
  - some set topics, no set sequence
  - interviewer can follow replies
  - less formal, for requirements gathering
- Semi-structured interview
  - set of questions available for interviewer to draw on if interviewee digresses or doesn't say much
- Prompted interview
  - draw out more information from interviewee, based on screen design or prototype → "what do you mean by…"

Questionnaires and surveys

- Focus is on preparation of unambiguous questions
- Pilot study important → once it's out there, it's final
- open questions:
  - respondent free to provide own answer
- closed questions:
  - respondent selects from set of alternative replies
  - usually some form of rating scale
  - Yes/No/maybe
  - Likert scale (strongly disagree to strongly agree)
  - Semantic difference (extremely | quite | slightly | neutral) with easy and difficult in two extremes
  - Rank order – rank 1-4 the following items in its ease of use

Which scale to use?

- Major task in measurement: systematically apply numbers to variables.
- Nominal (naming/category scale)
  - Differences between categories – qualitative.
  - represent categories where there is no basis for ordering the categories, e.g. male vs. female, ford vs. toyota.
- Ordinal (order):
  - involve categories that can be ordered along a pre-established dimension.
  - no way of knowing how different the categories are from one another, e.g. white, green, blue, brown belts.
- Ratio (numbers) :
  - Distance between adjacent numbers are equal.
  - Most ratio scales are counts of things (e.g. temperature).
  - There is reference to zero point.

Which scale to use (cont.)?

- Interval :
  - similar to standard numbering scales except that they do not have a true zero (distance between successive numbers is equal), e.g.: IQ (there is no 0).
- Why do we need to make the distinction?
  - It affects the statistical procedures that will be used in describing and analyzing data (parametric vs. non).
- Effective range of the scale
  - Every measure has an effective range for the population under study.
- Attenuation effect: if effective range is inadequate (distorts data & threatens the validity of the study).
  - Ceiling effect – restricted higher range
  - Floor effect - restricted lower range

Parametric vs. Non Parametric

- Mean, stdev vs. median, quartile
- Parametric statistics assumptions
  - Observations must be independent
  - Observations must be drawn from normally distributed populations
  - These populations must have the same variances
- Non-Parametric statistics assumptions
  - Observations are independent
  - Variable under study has underlying continuity
- What tests?
  1. Tests of differences between groups (independent samples)
  2. Tests of differences between variables (dependent samples)
  3. Tests of relationships between variables

Summary Table of Statistical Tests

Level of Measure-ment	Sample Characteristics					Correlation
	1 Sample	2 Sample		K Sample (i.e., >2)		
		Independent	Dependent	Independent	Dependent	
Categoric/ Nominal	X^2 or binom	X^2	Macnarmar's X^2	X^2	Cochran's Q	
Rank or Ordinal		Mann Whitney U	Matched Pairs Signed Ranks	Kruskal Wallis H	Friendman's ANOVA	Spearman's rho
Parametric (Interval & Ratio)	z test or t test	t test between groups	t test within groups	1 way ANOVA between groups	1W ANOVA (within or repeated measure)	Pearson's r
		Factorial (2 way) ANOVA				

A good survey

- Name your survey interestingly
- Write a short questionnaire
  - What is essential to know? What would be useful to know? What would be unnecessary?
- Use simple words
  - Don't: "What is the frequency of your automotive travel to your parents' residence in the last 30 days?"
  - Do: "About how many times have you driven to your parent's home in the last 30 days?"
- Relax your grammar
  - Don't make the questions sound too formal.
  - For example, the word "who" is appropriate in many instances when "whom" is technically correct.

A good survey

- Start with interesting questions
  - Start easy and attractive.
  - Put difficult or threatening questions down.
  - Voicing questions in the third person can be less threatening than questions voiced in the second person.
- Don't write leading/assuming questions
  - Don't ask "How many children do you have?"
- Assure a common understanding
  - Write in a way that everybody understand in the same way.
  - Don't assume that everyone has the same understanding of the facts or a common basis of knowledge.
  - Identify even commonly used abbreviations to be certain that everyone understands.

A good survey

- Balance rating scales
  - Mediate the scale so that there is room for both extremes.
- Don't make the list of choices too long
  - If the list of answer categories is long and unfamiliar, it is difficult for respondents to evaluate all of them. Keep the list of choices short.
- Avoid difficult concepts
  - Questions that involve difficult concepts cause variability due to understanding.
- Avoid difficult recall questions
  - People's memories are increasingly unreliable as you ask them to recall events farther and farther back in time.
- Avoid double negatives
- Think about order of questions

Contextual Inquiry

- Users and researchers participate to identify and understand usability problems within the normal working environment of the user
- Usability issues that go undetected in laboratory testing
  - Line counting in word processing
  - Unpacking and setting up equipment
- Differences from other methods include:
  - work context – larger tasks
  - time context – longer times
  - motivational context -- more user control
  - social context – social support included that is normally lacking in experiments
  - environment context – effects of artefacts

Participative Evaluation

- A.k.a. participatory design → Scandinavian idea
- Beyond standard user-centered design method
- Users considered domain experts and active participants
- cooperative prototyping, facilitated by
  - focus groups
  - designers work with users to prepare prototypes
  - iterative prototyping with user evaluation inseparable
  - tight feedback loop with designers
- But problems include:
  - may not get management commitment to let people go
  - union rep or users?
  - difficult to coordinate timing with various stakeholders

Predictive Evaluation

- Predict aspects of usage rather than observe and measure, usually by experts
- doesn't involve users → cheaper in time and efforts
- Inspection Methods
  - Standards inspections
  - Consistency inspection
  - Heuristic evaluation
  - "Discount" usability evaluation
  - Walkthroughs
- Modeling: The keystroke level model, GOMS (Goals-Operator-Method-Selection), ACT-R, EPIC (Executive Processes in Cognition) → in general cognitive architecture models

Inspection Methods

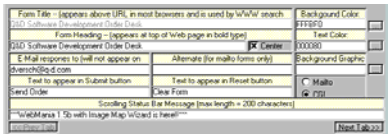
- Standards inspections
  - Experts inspect the interface for compliance with specified standards, e.g., WCAG
  - relatively little task knowledge required
- Consistency inspections
  - Teams of designers inspect a set of interfaces for a family of products
- Expert review
  - Experts simulate behavior of novices, experts and code breakers, try to anticipate problems
  - More efficient than user trials but not as good as the simulators
- Heuristics evaluation
  - Usage simulation in which system is evaluated against list of "heuristics"
  - Two passes per screen, flow from screen to screen

Nielsen's Heuristics

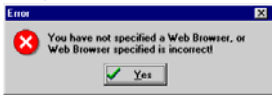
- Recognition rather than recall

- Flexibility and efficiency of use

  - Shortcut keys
- Aesthetic and minimalist design

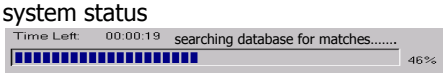


- Help users recognize, diagnose, and recover from errors
- Help and documentation



Nielsen's Heuristics

- Visibility of system status



- Match between system and the real world

- User control and freedom

- Consistency and standards

- Error prevention

Discount usability engineering

- Nielsen's heuristics: 5 independent experts produce 75% of usability problems
- Phase 1: usability testing + scenario construction (1-3 users)
- Phase 2: scenarios refined + heuristic evaluation
- "Discount" features
  - small scenarios, paper mockups
  - informal think-aloud (no psychologists)
  - Scenarios + think-aloud + heuristic evaluation
  - small number of heuristics
  - 2-3 testers sufficient

(Cognitive) Walkthrough

- A formalized way of imagining people's thoughts and actions when they use an interface for the first time.
- Variation: pluralistic walkthough (done in pair)
- Theory :
  - A user sets a goal to be accomplished by the system.
  - The user will search the system for the action that seems likely to make progress towards that goal.
  - By putting yourself in the user's shoes you can figure out where the problem might occur.
- Must be complemented with user studies because:
  - Is diagnostic, not prescriptive
  - Focuses mostly on novice users
  - Relies on the ability of evaluator to put themselves in the users shoes

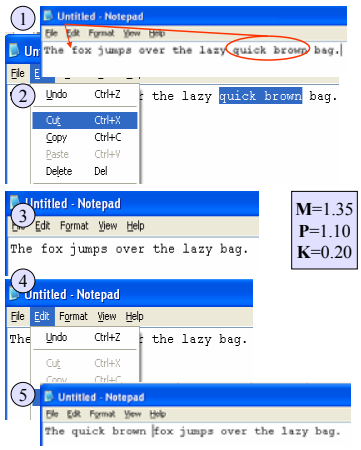
Cognitive Walkthrough How To

- Begin by collecting:
  - An idea of who the users will be and their characteristics
  - Task description
  - Description of the interface (a paper prototype)
  - Written list of the actions to complete the task given the interface (scenario)
- For each action in the sequence (scenario)
  - tell the story of why the user will do it
  - ask critical questions:
  1. will the user be trying to produce the effect?
  2. will the user see the correct control?
  3. will the user see that the control produces the desired effect?
  4. will the user understand the feedback to proceed?
  5. will the user select a different control instead?

Modeling: keystroke level model

- Goal: calculate task performance times for expert users (no error)
- Requires
  - specification of system functionality
  - task analysis, breakdown of each task into its components
- Time to execute sum of:
  - Tk - keystroking (0.35 sec)
  - Tp - pointing (1.10)
  - Td - drawing (problem-dependent)
  - Tm - mental (1.35)
  - Th - homing (0.4)
  - Tr - system response (1.2)

Method Used
Cut-and-paste-using-menus



M=1.35
P=1.10
K=0.20

Description	Opr	Dur (s)
Mentally Prepare	M	1.35
Move cursor to "quick"	P	1.10
Double-click mouse button	K	0.40
Move cursor to "brown"	P	1.10
Shift-click mouse button	K	0.40
Mentally Prepare	M	1.35
Move cursor to Edit menu	P	1.10
Click mouse button	K	0.20
Move cursor to Cut menu	P	1.10
Click mouse button	K	0.20
Mentally Prepare	M	1.35
Move cursor to before "fox"	P	1.10
Click mouse button	K	0.20
Mentally Prepare	M	1.35
Move cursor to Edit menu	P	1.10
Click mouse button	K	0.20
Move cursor to Paste menu item	P	1.10
Click mouse button	K	0.20
TOTAL PREDICTED TIME		14.90

Controlled Experiments

- Typically narrowly defined, evaluate particular aspects such as:
  - menu depth vs. breadth
  - new vs. old icon design
- Major issues:
  - What to change? What to keep constant? What to measure?
  - Hypothesis, stated in a way that can be tested.
  - Statistical tests: which ones, why?
  - Uncontrolled items (disruption, order effect, fatigue)
- Variables
  - Independent: the manipulated one (the icon designs)
  - Dependent: the measured one (time, error)
  - Control: age, gender, expertise

Psychometrics Testing

- Psychometrics = scientific measurement of individual differences (personality and intelligence)
  - Psychological qualities of individuals
  - Make predictions about behavior
- Psychometric test: psychology = microscope: biology (Dawis, 1992)
- Test = an objective, systematic and standardized measure of a sample of behaviour
  - Objective = every observer of an event would produce an identical account of what took place
  - Systematic = a methodical and consistent approach to understanding an event
  - Standardized = observations of an event are made in a prescribed manner

Psychometrics Testing

- Types:
  - Cognitive ability: measures an individual's ability to process information from their environment
  - Personality measures: people's dispositions to behave in certain ways in certain situations
- Categories:
  - Normative tests – most psychometric tests where data exists which tell us the range of scores expected from the population under consideration e.g. IQ scores
  - Criterion referenced tests – commonly used in education where a candidate has to meet some pre-arranged standard.
  - Idiographic tests – used in therapy to observe an individual's progress over time

Personality measure

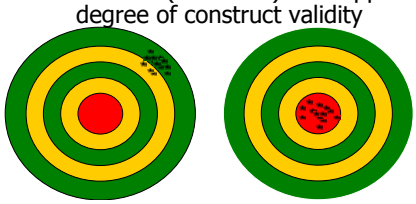
- Used by:
  - Criminal psychologists to measure impulsivity and its relation to crime
  - Health psychologists to measure people's optimism in relation to their response to cancer diagnosis
  - Occupational psychologists to predict job performance and job suitability.
- Forms:
  - Objective tests: Individuals are asked to rate their own actions or feelings in set situations
  - Projective tests: Individuals are asked to formulate an unstructured response to some form of ambiguous stimuli e.g. Rorschach ink-blot test (Rorschach, 1921)

Three concepts for a good test

- Standardisation: ensures that the conditions are as similar as possible for all individuals who are given the test.
- Reliability – a test must measure the same thing in the same way every time someone takes it
  - Internal consistency – all the parts of the test are reliable throughout
  - Test–retest – the test remains valid over time
  - Inter-rater – independent observers rate the same sample should produce more or less the same results
- Validity: a test must provide an indication of the strength of our conclusions, inferences or propositions

Test Validity

- There are four types of test validity:
  - Face validity: does your test appear to measure what it purports to measure
  - Concurrent validity: does your test of honesty correlate with existing standardised tests of honesty
  - Predictive validity: do the results of your test predict future behaviour
  - Construct validity: if all our hypotheses about the test variable (construct) are supported then we have a high degree of construct validity



Problems Affecting Results

- Social Desirability – people feel they are being judged and so alter their answers accordingly for 2 reasons:
  - Self-deception – individuals are overly optimistic in their perceptions of their own positive personality features and play down their perceived negative aspects
  - Impression management – individuals try to appear 'nice' because they fear social disapproval
- Mood
  - people in a good mood might answer the questionnaire completely differently than if they were in a bad mood
- Features of the environment (noise, heat & light)
  - Hancock (1986) has shown that high temperature has a significant negative effect on vigilance, attention, memory and reaction time

Harder Problems

- Ecological Validity
  - Ensure test what an individual do in their daily life, not in a research environment
  - The tension between controlled experiment and in-context study
- Cultural bias
  - Most standardized psychometric tests are based on western definitions and western cultural practices
  - the possibility of bias in such tests against members of ethnic subgroups of the population
  - e.g. newly arrived refugees will have difficulty with an intelligence test which asks them to name leaders of the country to which they escaped from

Culture-Free Tests

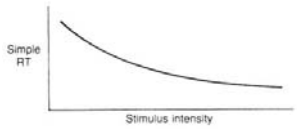
- Attempts have been made to develop culture-free tests of intelligence, but on the whole these attempts have not been successful, because:
  - Conceptions of intelligence vary widely from culture to culture
  - even if the content of a test can be made culture-free, culture itself will still affect the results through directing attitudes towards tests, test-taking, competition, etc
- Some untimed and language-free (almost) tests
  - The Leiter International Performance Scale – Revised (Roid & Miller, 1997): 4 domains of functioning: reasoning, visualisation, attention and memory

http://www.stoeltingco.com/tests/downloads/subtest%20webclip.mov

  - The Ravens Progressive Matrices (Court & Ravens, 1995): general cognitive ability

Performance Test: SRT

- Simple Reaction Time (motor speed test)
- No uncertainty what the signal is, and how to respond
- Affected by:
  - Stimulus Modality: RT(aud) < RT(vis)
  - Stimulus Intensity
    - More intense stimuli lead to shorter RTs
    - Can be modeled using SDT (aggregation of neural evidence over time)
    - Can raise or lower criterion (e.g., false start for sprinter)
  - Temporal Uncertainty
    - Greater uncertainty increases RT

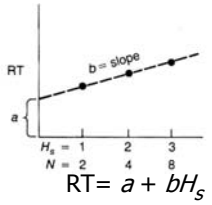


Performance Test: CRT

- Choice Reaction Time (motor + cognitive)
- there can be more than one signal, and more than one type of response
- Factors affecting SRT also affect CRT
- In a choice response time situation, the subject is transmitting information from stimulus to response in the information theory sense
- Hyman's (1953) experiment: S observes set of lights; responds with particular response when particular light flashes
- *Hick-Hyman Law* (H-H Law)
  - Choice RT increases linearly with stimulus information
  - RT= $a + bH_s$

Hick-Hyman Law

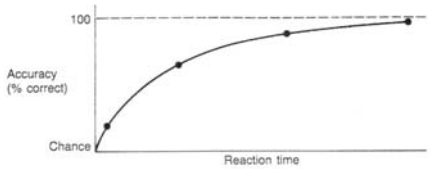
- Slope *b* is about 170 ms/bit—amount of extra time resulting from each added bit of stimulus information to be processed
- Can derive *information transmission rate* (bandwidth) by $1/b$ = 0.00588 bits/ms = 5.88 bits/s
- Intercept *a* (around 180 ms) represents time to encode the stimulus and execute the response) – factors unrelated to the stimulus information
- Doesn't matter how the amount of information in the stimuli is varied
- Affected by number of alternatives, probability, context
- Tested many times with different kinds of stimuli and responses, and is generally accurate



RT= $a + bH_s$

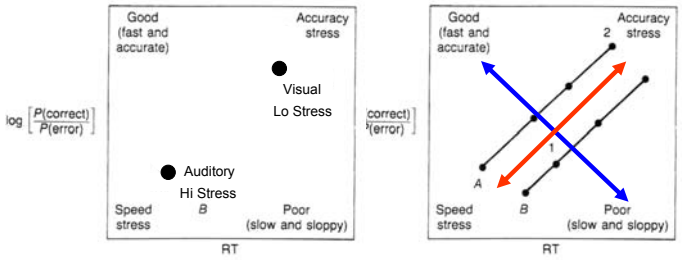
Speed-Accuracy Tradeoff

- People tend to make more errors when they respond more rapidly; if they take longer they tend to be more accurate
- When you push people to be extremely accurate, reaction time increases a lot for little increase in accuracy—diminishing returns
- To get most efficient performance:
  - For easy task, best to emphasize bandwidth
  - For hard task, best to emphasize speed



SAOC (Speed-Accuracy Op Char)

- "NW is best; SE is least" good vs. poor performance
- Going from southwest to northeast—moving along the SAOC—represents different speed-accuracy tradeoff settings
- Auditory processing/hi stress tends to lead to more rapid, error-prone performance (quick and dirty) than does visual processing/lo stress

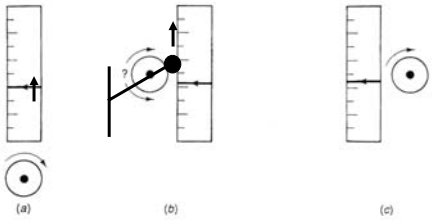


Compatibility

- Stimulus-Response Compatibility (natural mapping)
  - between displayed information and response or control
  - *Static sense*: Compatibility between a display location and the location of the response
  - *Dynamic sense*: Compatibility between display movement and movement involved in the response
  - High SRC = fewer mental operations, quicker response
- Locational Compatibility
  - We have natural tendency to move or orient towards source of stimulation in environment—infants will orient to new pictures, new faces
  - So why not put the control and the display in the same location? *Colocation principle*
  - If can not, use congruence principle

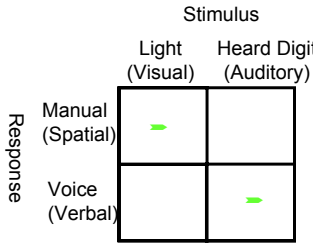
Movement Compatibility

- Typically movement of the control should correspond to the movement in the display
- When not possible, there are common conventions to map display and control movements
- Movement proximity → placing control next to object



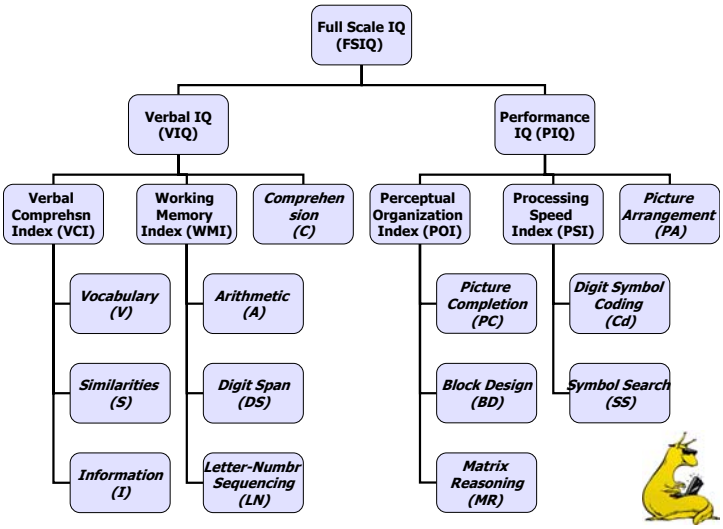
Modality Compatibility

- S-R compatibility can be affected by stimulus and response modality as well as by spatial correspondence
- If stimulus is a light, faster CRT for a manual response than for a voice response
- If stimulus is a heard digit, faster with naming response than with a spatial pointing response



Intelligence Test: WAIS

- Intelligence
  - the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment.
  - Global because it characterizes individual's behavior as a whole
  - Aggregate because it is composed of elements or abilities that are qualitatively differentiable
- 3 intelligence test, individually administered
  - The first was the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939)
  - Wechsler Pre-school and Primary Scale of Intelligence (WPPSI) - 3-7 years
  - Wechsler Intelligence scale for Children (WISC) - 7-16
  - Wechsler Adult Intelligence Scale (WAIS) – 16-89



What are they measuring?

- VCI - Verbal Comprehension - emphasis on crystallized intelligence (knowledge application)
- PRI - Perceptual Organization - emphasis on fluid intelligence (new learning)
- WMI - Working Memory - emphasis on short-term memory/retrieval (auditory)
- PSI - Processing Speed - emphasis on mental quickness (task performance with focused concentration & attention)
- Wechsler's tests provide three scores:
  1. a verbal IQ (VIQ)
  2. a performance IQ (PIQ)
  3. a composite, single full-scale IQ score based on the combined scores