

Tests of Functional Vision in CVI: Evidence Based Approaches

1:30-3pm

Corinna Bauer

MASSACHUSETTS SCHEPENS EYE
EYE AND EAR RESEARCH INSTITUTE

HARVARD MEDICAL SCHOOL DEPARTMENT OF
Ophthalmology

What Makes a “Good” Assessment?

- Measures what it is supposed to
- Doesn't measures what it says it doesn't
- Repeatable
- Same result regardless of who gives the assessment (consistent)
- Other factors??

COSMIN Standards

Consensus-based Standards for the selection of health status
Measurement Instruments

COSMIN standards



- Internal Consistency – Does the test measure what it is meant to?
- Reliability – Does the test provide stable, consistent results?
 - Over time, between administrators, across items in the test
- Measurement Error – Difference between the measured value and its true value
- Content Validity – How well the test measures the behavior for which it is intended
 - Correspondence between test items and symptoms
- Structural Validity – How well the questionnaire score reflects the construct being measured
- Hypothesis Testing – Test the size and direction of effect in a random sample
- Cross-cultural Validity – Can the test be used in different populations and cultural groups
- Criterion Validity – How well the tool measurements compare to the 'gold standard'
- Responsiveness – The ability of the tool to detect change over time in the measured construct

BMC Med Res Methodol. 2010;10:22

Reliability

- "The degree to which the result of a measurement, calculation, or specification can be depended on to be accurate."
- Internal Reliability
 - Consistency of results across items within a test
- External Reliability
 - Extent to which a measure varies from one use to another
 - Test re-test: stability of test over time
 - Inter-rater: consistent results across different independent raters
 - Inter-method reliability: consistent results with other tests/instruments used
- Reliability does not equal validity

How to Evaluate Reliability

- Internal reliability/consistency – Cronbach's alpha
 - Used to determine if the designed test accurately measures the variable of interest
- Reliability (correlation) coefficient – correlation between 2 or more variables
 - 0.9-1.0 – excellent reliability
 - 0.8-0.9 – good reliability
 - 0.7-0.8 – acceptable reliability
 - 0.6-0.7 – questionable reliability
- Test-re-test – correlation between two or more separate occasions
- Inter-rater – correlation between scores from two or more administrators
 - Note: high correlation does not ensure the test is administered correctly, only that the test is being measured the same
- The more items in the test, the larger the sample needed

Validity

- “The extent to which a test accurately measures what it is supposed to measure”
- Content Validity
 - Measures what it was designed to measure
 - E.g. a comprehensive math test would lack content validity if good scores depended mainly on knowledge of English or if it only contained algebra questions
- Face Validity
 - Ability of an instrument to be understandable and relevant for the targeted population
- Criterion-related Validity
 - Useful for predicting a person's performance on an external criterion measure
 - E.g. score on test A predicts scores in test B
- Construct Validity
 - The degree to which a test measures what it claims to be measuring
 - The appropriateness of the inferences made on the basis of observations or measurements
 - Content Validity – The degree to which an assessment is relevant to or representative of the targeted construct it is designed to measure
 - Enables instrument to make meaningful and appropriate inferences/decisions based on the scores
 - Concurrent Validity - Measures how consistent the results are compared to other already validated tests
 - Predictive Validity - The extent to which test predicts score on some criterion measure

How do we Determine Validity

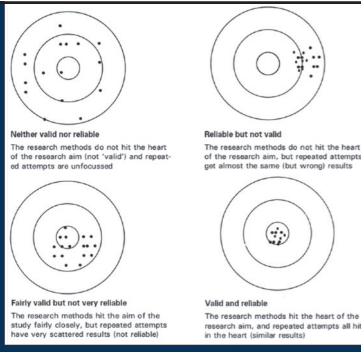
- “The extent to which a test accurately measures what it is supposed to measure”
- Concurrent Validity: Correlation with existing test measuring same construct
- Content Validity: Expert Opinion, Cronbach's alpha
- Predictive Validity: Correlation with predicted outcome
- Need to make sure that you have a large enough sample

Relationship between Validity and Reliability

- If a test is unreliable, it cannot be valid
- For a test to be valid, it must be reliable
- Just because a test is reliable, does not mean it will be valid
- Reliability is a necessary but not sufficient condition for validity
- High alpha does not guarantee construct of interest is being measured or that important concepts are not missing
- High test-retest reliability does not imply that all items are relevant or that important concepts are not missing

Examples:

- Every day for the past year, your scale at home says that you weigh 200 lbs. At the doctor's office, you weigh 150 lbs.
- Is your home scale reliable?
- Is your home scale valid?



Examples:

- You and a colleague are testing a new device to measure temperature. You each take temperatures with the new device and the "gold standard" thermometer.
- You get different values than your colleague and the values you obtain change on the same person between week 1 and week 2
- Does the new test have good external reliability? Internal reliability? Test-retest reliability? Inter-rater reliability?
- Is this new measurement for temperature valid?

Reliability and Validity: What's the Big Deal?

- Validity is important because if the test does not measure what it is intended to, then the results cannot be used to answer that question
 - The results cannot be used to generalize the findings
 - Ensures that results can be used effectively
- Reliability is important because if the results are inconsistent across time or administrators, then changes in the test result cannot be trusted – uncertain if change in outcome due to test or actual change

Sensitivity & Specificity

- Measures of classification accuracy
- Sensitivity: True positive rate
 - Proportion of actual positive that are correctly identified as such
- Specificity: True negative rate
 - Proportion of actual negatives that are correctly identified as such
- Positive Predictive Value
 - Percentage of patients with a positive test who actually have the condition
- Negative Predictive Value
 - Percentage of patients with a negative test who do not have the condition

Evaluating Sensitivity & Specificity

- Want high sensitivity and high specificity
- Sensitivity = true positive / (true positive + false negative)
- Specificity = true negative / (true negative + false positive)
- PPV = true positive / (true positive + false positive)
- NPV = true negative / (false negative + true negative)

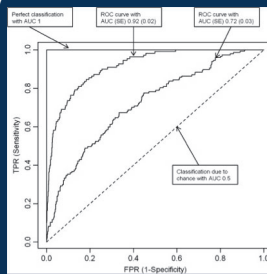
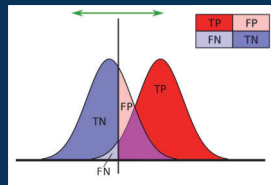
Gold standard disease present		Gold standard disease absent		Total test positives	
Test positive	a	b	a+b		
Test negative	c	d	c+d		
	a+c	b+d	a+b+c+d		

[Indian J Ophthalmol. 2008 Jan-Feb; 56\(1\): 45-50.](#)

Disease present		Disease absent	
Test positive	a (TP)	b (FP)	
Test negative	c (FN)	d (TN)	
	a+c	b+d	

TP: True positive, FP: False positive, FN: False negative, TN: True negative

Sensitivity and Specificity



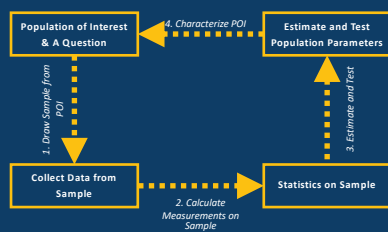
Sensitivity and Specificity: Why Do They Matter?

- If a test claims 100% sensitivity, it may not be very specific and provide a large number of false positives
- Tests and screening tools should be as accurate as possible, particularly when it governs who gets intervention or treatment

The Importance of Good Study Design

Factors to Keep in Mind in your Own Studies

Process of Research Inference

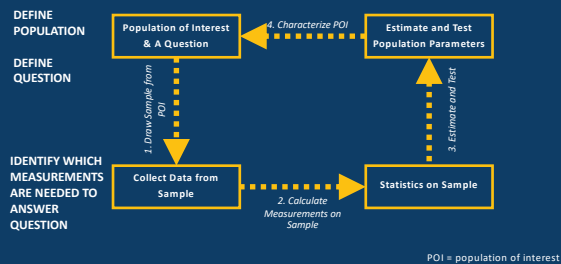


POI = population of interest

Study Design Basics

- Multiple types of study designs
 - All aimed at answering some sort of specific topic or question
- Study design will depend on the question you want to ask and the type of data you will be collecting or have already
 - Question needs to be clearly defined
 - Outcome needs to be measurable
- Each type of study design has strengths and limitations
 - Vary in terms of bias and overall strength

Study Design

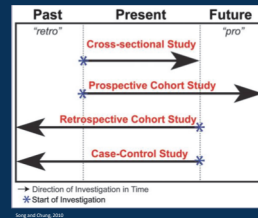


Types of Study Design

- Classified broadly based on two characteristics:
 1. Time period in which the data is collected
 2. Employ an experimental treatment

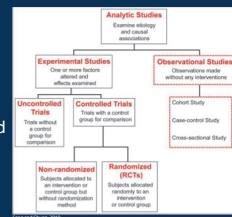
Study Design

- Prospective – subjects recruited and data collected on subsequent events
- Retrospective – subjects recruited and data is collected on prior events or exposures
- Cross-Sectional – subjects recruited and data collected at a single fixed time point, rather than multiple time points



Study Design

- Experimental – impose a treatment/intervention and collect information on responses.
 - Aims to determine if treatment affects response
 - Can only be done prospectively
- Observational – observe the individuals and collect variables of interest
 - Do not impose any condition on the subjects
 - Can be prospective, retrospective, or cross-sectional



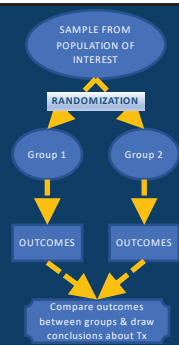
Key Concepts

"Any proof of effectiveness cannot be obtained from a non-randomized study"
Norum et al., 2012

- Bias – any influence that acts to make the observed results not actually representative of the true results. Can occur in the design, conduct, or interpretation of the study.
 - E.g. In an study involving patient interviews, could bias the results by asking leading questions or changes in tone of voice that might influence the response
 - "Do you have difficulty finding your room in the supermarket?" vs.
 - "How sometimes have difficulty finding your room in the supermarket, don't you?"
- Variability – spread of outcomes between or within individuals
 - High variability can make it difficult to draw conclusions because it can mask the differences between groups
- Randomization – participants are randomly assigned to groups
 - E.g. heads or tails
- Blinding – mask the identity of the assigned group/intervention
 - Helps avoid bias
 - E.g. in a clinical trial, usually the patient and the physician are not aware which group that individual is assigned to
- Placebo effect – beneficial response produced by a placebo that are due to the patient's belief in the treatment, rather than the treatment/placebo itself
 - E.g. controls for Drug A given a sugar pill, but still improve in symptoms

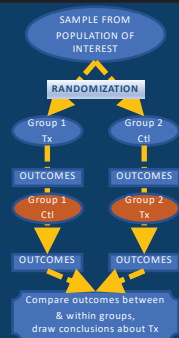
Experimental Study: Randomized Control Trial (RCT)

- Gold Standard in study design
 - Reduces bias (affects accuracy)
 - Reduces variability (affects precision)
- Sample from population is **randomly assigned** to groups
 - e.g. treatment and control, treatment 1 and 2, etc.
- Randomization is very important!
 - Helps ensure that potentially confounding effects are balanced between the groups
 - Can more safely assume that difference in outcomes attributed to treatment
 - Enables cause and effect relationship to be drawn
- Prospective, randomized, and comparison group



Experimental Study: Cross-Over Design

- In Cross-Over Design, each subject receives all treatments
 - Enables comparisons within and between groups
 - Subject can serve as their own control because they are part of *both* groups
- Repeated measures: measure outcomes before and after the switchover
- Sometimes unethical for one group not to receive the treatment
- *OR* want each subject to serve as their own control



Observational Study Types

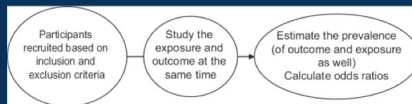
- Ecological
- Cross-Sectional
- Cohort
- Case-control

Observational Study Types: Ecological Design

- At least 1 variable measured at a population or group level, rather than at the individual level
 - E.g. Relationship between per capita preterm birth and prevalence of visual impairments across regions in the USA
 - Per capita preterm birth: ecological variable
 - Cases of visual impairment per 1,000 within the region is group data because region, rather than individual information
- Often used to monitor population health
- can make large-scale comparisons (e.g. between countries, states, provinces)
- Can study relationship between population-level exposure to risk factors and disease

Observational Study Types: Cross-sectional Design

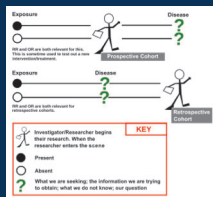
- Measure outcome and the exposure in the participants at the same time at a single time point
 - Assessing prevalence of condition
 - Population-based surveys
 - Calculate odds ratios (e.g. males have a higher odds of having a beard)



Smith, 2014

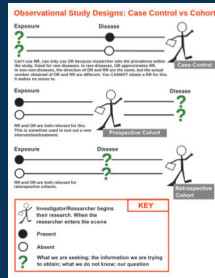
Observational Study Types: Cohort Design

- Evaluate the effect of exposure (e.g. exposure to radiation, presence of a genetic factor, etc.) on the outcome of interest (e.g. incidence of cancer)
- Prospective: Recruit subjects and collect baseline exposure data *before* any subjects have developed the outcome of interest
 - E.g. Framingham Heart Study, Nurses Health Study
- Retrospective: Recruit subjects and collect data after exposure and compare outcome of interest
 - E.g. Comparing the incidence of cancer in the Smith family between those with and without genetic factor



Observational Study Types: Case-Control Design

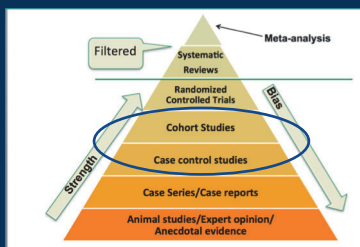
- Compare a group of subjects with condition and those without the condition and compare the level of exposure to factors of interest
 - Suggests an association between the level of exposure and the condition of interest
- E.g. Recruit group of patients with cancer and those without and evaluate differences in lifestyle factors between the groups that may be associated with cancer (e.g. smoking)



Summary

Design	Type	Strengths	Limitations
Randomized Control Trial	Experimental, Prospective	Randomized Comparison Group, can apply different treatment designs	Can be time consuming and expensive, not ethical for all situations
Ecological	Observational	Useful for hypothesis generation	Findings for group may not apply to individual
Cross-Sectional	Observational	Faster and less expensive than cohort studies	Cannot determine trends over time
Cohort	Observational, Prospective	Can assess temporal changes	Difficult for rare diseases or long latency periods
Case-Control	Observational, Retrospective	Useful for rare outcomes	Could be recall or misclassification bias

Study Design Overview



Summary of Things to Keep in Mind When Evaluating the Results of Studies

- Association is not causation
- Importance of reproducibility
- Isolating effects with control and experimental groups
- Learning effects (doing anything twice will likely make you better)
- Unbiased, "blinded" assessment – staying objective
- The power of a big 'n' (large sample sizes matter)

Tools for CVI

Applying a Critical Eye to CVI: Questions to Ask Yourself When Evaluating a Study or New Tool

Does this study use:

- Treatment AND control groups?
- Controlled interventions?
- Adequate statistical analysis?
- Evidence-based approaches?
- Clear and measurable outcomes?

Does this tool have

- Clear and measurable outcomes?
- Evidence-based design?
- Reliability?
 - Are the results consistent over time, between administrators, or sites?
- Content Validity?
 - Does it measure what it is intended to?
- Structural Validity?
 - Do the questions reflect the construct being measured?
- Cross-cultural validity?
 - Is it appropriate to use in my population?
- Criterion validity?
 - How does it compare against the "gold standard"?
- Responsiveness
 - Is it able to detect meaningful change over time without influence of bias?
- Can it be used to test a hypothesis?

CVIT 3-6: Children's Visual Impairment Test for 3-6 y/o

- 14 subtests across 4 domains of visual perception
 - Object Recognition
 - Degraded Object Recognition
 - Motion Perception
 - Global-Local Processing

Developmental, Behavioral & Clinical Research (Original Article)

Assessment tool for visual perception deficits in cerebral visual impairment: development and normative data of typically developing children

KATHLEEN VANCLEEF^{1,2} | EVA JANSSEN^{1,2} | NICHOLAS PETER¹ | JOHAN WAGNER^{1,2} | ELS DIERICK^{1,2}

Developmental, Behavioral & Clinical Research (Original Article)

Assessment tool for visual perception deficits in cerebral visual impairment: reliability and validity

KATHLEEN VANCLEEF^{1,2} | EVA JANSSEN^{1,2} | NICHOLAS PETER¹ | JOHAN WAGNER^{1,2} | ELS DIERICK^{1,2}

Figure 3: Boxplots of Children's Visual Impairment Test for 3- to 6-year-olds (CVIT 3-6) scores for our four validation groups: children with cerebral visual impairment (CVI), intellectual impairment (II), typically developing (TO), children with acquired non-specific visual loss (VA).

The test

CVIT 3-6

Designed for developmental age range 3-6 years and for children with visual acuity > 0,2

Object Recognition

Degraded Object Recognition

Motion Perception

Global-Local Processing

Slide courtesy of Els Dierick CVI Minisymposium Orribus Vancleef et al, DMCK, accepted

Reliability

N = 300, age 3-6, Cronbach's $\alpha = .65$
 → Internal consistency is rather low as expected

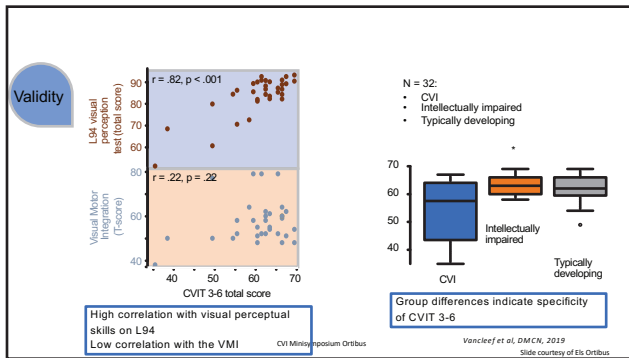
Test-retest reliability

N = 32
 2-3 weeks interval

Small learning effect
 $\Delta = -1.8, p = .01$

$r = .86, p < .001$
 → Very good reliability

Slide courtesy of Els Dierick CVI Minisymposium Orribus Vancleef et al, DMCK, accepted



CVIT 3-6 summary

- The CVIT-3-6 is a scientifically sound tool for identifying visual perceptual/ 'higher vision' difficulties in children with a developmental age 3- 6 years;
- It does not involve motor skills and could be used with a wide group of children including those with cerebral palsy
- It is an online tool, fun and easy to use in the clinic (<https://psytests.be/clinicians/test-centrum/cvi-t.php>)
- It potentially assesses and identifies children who would be in Sakki et al Cluster subtype A1 and Cluster subtype A2

Slide courtesy of Els Orribus

CVI Mini-symposium Orribus

Teach CVI



- teachCVI is a European partnership that aims to create collaborative tools for teachers and healthcare professionals that will help bridge the gap between teachers/educators and healthcare professionals so that they can work together to benefit the target group: children with CVI

- **Aims of the project:**
- Making a tool for healthcare professionals and educators to screen for CVI
- Creating a common database of tools for CVI detection
- Producing resources for teachers to support their work in the assessment of CVI
- Making teaching methodologies to enable the child's access to literacy, this includes training and teaching materials for teachers/educators of children with CVI
- <https://www.teachcvi.net>
- Series of assessments, screening tools, and other helpful resources

L-94

Assessment of cerebral visual impairment with the LSA visual perceptual battery: clinical value and correlation with MRI findings

• Perceptual battery created in Leuven, Belgium in 1994

• Shows correlation with MRI and agreement with other tests of visual perceptual functions

Visual Matching (VISM)

Target drawing Matching alternatives

Overlapping Line Drawings (OVER)

Complete overlap Partial overlap Touching No overlap

Die View Task (DE VOT)

Unconventional Object Views (VIEW)

Foreshortening of horizontal axis Canonical viewpoint

Line Drawings Occluded by Noise (NOLSE)

60% occluded 40% occluded 20% occluded

Infant Screening Tools - Italy

Neonatal Assessment Visual European Grid (NAVEG): Unveiling neurological risk

Andrea Ruffini, Marco Girelli, Cristina Montemali, Cristina Chiodi, Laura Molteni, MD, Lutz B. Meubert, NAVEG Study Group, Elina Pavesi

Ocular Visual Component:

- ✓ Red Reflex
- ✓ Photomotor Reflex

Motor Visual Component:

- ✓ Fixation
- ✓ Horizontal Smooth Pursuit
- ✓ Vertical Smooth Pursuit
- ✓ Saccadic movements

Basic Visual Functions:

- ✓ Contrast sensitivity
- ✓ Visual acuity
- ✓ Visual field
- ✓ Stereopsis
- ✓ Depth perception

Item's score:

- 0 normal
- 1 borderline
- 2 impaired
- NA not assessable

Screening tools for VI in schools - UK

• Non-verbal children:

- Lea-Grime
- Cardiff Cards
- Puppet Face
- Hidden Image (contrast)

• Young Children (ability to match)

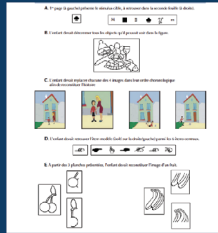
- BUST (near and distance)
- Key Picture test
- Donkain-Shire Visual Acuity System
- Colour Vision Made Easy
- PV 16 Colour Test
- Lea Symbols/High/Low Contrast test

• Older Children

- MacLure Reading Type for Children (near)
- Kays Crowded LogMAR letter test
- Snellen Chart
- Pelli Robson Chart (Contrast Sensitivity)
- Precision Vision Colour Test as above
- Ishihara

Varies Across districts and countries

Screening tools for VI in schools - France



Vilayphonh et al., 2009

Visual Skills Inventory: 4-8, 9-12 y/o

- Structured clinical history taking inventory
- Home and School strategies
- <https://www.ulster.ac.uk/research/topic/biomedical-sciences/research/optometry-and-vision-science-research-group/vision-resources/resources-for-professionals/cerebral-visual-impairment-assessment>

How can we increase our knowledge of CVIs to improve patient care?

- Rigorous scientific evaluations
- Large studies with diverse populations or multiple replications
- Significant and sustained effects

Evidence-Based Principles

- "A systematic approach to clinical problem solving which allows the integration of the best available research evidence with clinical expertise and patient values"
• *Masic et al., 2008*
- "we would be better to base our decisions on the collective experience of thousands of clinicians treating millions of patients, rather than on what individuals have seen and felt"
• *Greenhalgh*

The Big Picture

Use neuroscience to determine the anatomical substrate of visual dysfunctions associated with CVI. In turn, this can help inform the definition, diagnosis, and treatment of CVIs.

- For example, CVI initially defined mainly on acuity and visual field loss and from "damage to the visual system between the lateral geniculate nucleus and the cortex"
• *Steinadam 1989*

We now know that CVIs extend to include many perceptual functions and that it can be caused by injury to other parts of the brain involved with visual processing, not just to the primary visual pathway.

- "A verifiable visual dysfunction which cannot be attributed to disorders of the anterior visual pathways or any potentially co-occurring ocular impairment"
• *Sakki et al., 2018*

Is there consensus in defining childhood cerebral visual impairment? A systematic review of terminology and definitions
Hanna E.A. Sakki,¹ Naomi J. Dale,^{1,2} Jennifer Sargent,² Teresa Perez-Rocha,³ Richard Bowman^{1,4}

"The benefits of evidence-based medicine, when properly applied, are obvious. We can use test characteristics and results to make better diagnoses. We can use evidence from treatments to help people make better choices once diagnoses are made. We can devise research to give us the information we are lacking to improve lives. And, when we have enough studies available, we can look at them together to make widespread recommendations with more confidence than we'd otherwise be able to."

Aaron E. Carroll
TheUpshot, The New York Times
Dec. 27, 2017