



PEARL

ISCEV and IPS guideline for the full-field stimulus test (FST)

Jolly, J. K.; Grigg, J. R.; McKendrick, A. M.; Fujinami, K.; Cideciyan, A. V.; Thompson, D. A.; Matsumoto, C.; Asaoka, R.; Johnson, C.; Dul, M. W.; Artes, P. H.; Robson, A. G.

Published in:

Documenta Ophthalmologica

DOI:

[10.1007/s10633-023-09962-7](https://doi.org/10.1007/s10633-023-09962-7)

Publication date:

2024

Document version:

Publisher's PDF, also known as Version of record

Link:

[Link to publication in PEARL](#)

Citation for published version (APA):

Jolly, J. K., Grigg, J. R., McKendrick, A. M., Fujinami, K., Cideciyan, A. V., Thompson, D. A., Matsumoto, C., Asaoka, R., Johnson, C., Dul, M. W., Artes, P. H., & Robson, A. G. (2024). ISCEV and IPS guideline for the full-field stimulus test (FST). *Documenta Ophthalmologica*, 148(1), 3-14. <https://doi.org/10.1007/s10633-023-09962-7>

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Wherever possible please cite the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.



ISCEV and IPS guideline for the full-field stimulus test (FST)

J. K. Jolly · J. R. Grigg · A. M. McKendrick · K. Fujinami ·
A. V. Cideciyan · D. A. Thompson · C. Matsumoto · R. Asaoka ·
C. Johnson · M. W. Dul · P. H. Artes · A. G. Robson

Received: 11 December 2023 / Accepted: 15 December 2023
© The Author(s) 2024

Abstract The full-field stimulus test (FST) is a psychophysical technique designed for the measurement of visual function in low vision. The method involves the use of a ganzfeld stimulator, as used in routine full-field electroretinography, to deliver full-field flashes of light. This guideline was developed jointly by the International Society for Clinical Electrophysiology of Vision (ISCEV) and Imaging and Perimetry Society (IPS) in order to provide technical

information, promote consistency of testing and reporting, and encourage convergence of methods for FST. It is intended to aid practitioners and guide the formulation of FST protocols, with a view to future standardisation.

Keywords Threshold · Sensitivity · Scotopic · Photopic · Clinical standards · International Society for Clinical Electrophysiology of Vision · Imaging and Perimetry Society

J. K. Jolly, J. R. Grigg, K. Fujinami, A. V. Cideciyan, D. A. Thompson, A. G. Robson representing ISCEV.

A. M. McKendrick, C. Matsumoto, R. Asaoka, C. Johnson, M. W. Dul, P. H. Artes representing IPS.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10633-023-09962-7>.

J. K. Jolly (✉)
Vision and Eye Research Institute, Anglia Ruskin University, Young Street, Cambridge CB1 2LZ, UK
e-mail: info@jvisionsci.com

J. R. Grigg
Save Sight Institute, Specialty of Clinical Ophthalmology and Eye Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

J. R. Grigg
Eye Genetics Research Unit, Sydney Children's Hospitals Network, Save Sight Institute, Children's Medical Research Institute, The University of Sydney, Sydney, NSW, Australia

Introduction

The full-field stimulus test (FST) was developed by Roman and colleagues for the assessment of vision in patients with severe vision loss. Roman et al. 2005 [1] used a modified perimeter to introduce the concept of

A. M. McKendrick
Lions Eye Institute, University of Western Australia, Perth, Australia

A. M. McKendrick
School of Allied Health, University of Western Australia, Crawley, Australia

K. Fujinami
Laboratory of Visual Physiology, National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Center, Tokyo, Japan

perception with full-field flash stimuli using manufacturer's software. Subsequently, Roman and colleagues used a Ganzfeld stimulator, as used for full-field electroretinography (ERG), to extend the dynamic range of available stimuli, with custom software to drive the system [2]. For a pre-defined stimulus and adaptation state, FST is used to provide a measure of visual function originating from any location in the retina and is presumed to originate from the most sensitive photoreceptors [1, 3–5].

The technique has been used increasingly in clinical trials for novel therapies, particularly gene therapy, to assess the restoration or preservation of retinal function [6]. FST is especially useful for detecting residual vision in patients with severe vision loss, including those with undetectable or severely abnormal full-field ERGs, or when poor fixation or nystagmus makes visual field tests difficult or impossible to perform.

The International Society for Clinical Electrophysiology of Vision (ISCEV, www.iscev.org) publishes standards, guidelines and extended protocols for electrophysiological methods including the full-field ERG [7–21]; the Imaging and Perimetry Society

(IPS, www.perimetry.org) publishes guidelines for psychophysical tests such as perimetry [22]. The FST and full-field ERG require diffuse flashes of light and are frequently performed using the same Ganzfeld flash stimulator. This document is a collaboration between the ISCEV and the IPS and is a guideline for FST testing, informed by methods that have been published and graded for quality. It is intended to aid practitioners and guide the formulation of FST protocols, to promote conformity and to facilitate meaningful inter-laboratory and inter-study comparisons, with a view to future standardization of routinely used FST methods.

It is highlighted that the term “FST” has been given a number of definitions historically in the literature and that this guideline defines FST as full-field stimulus test. Online Appendix 1 provides definitions of other relevant terminology.

Scope and applications

The FST has value to test and monitor disease progression and/or treatment efficacy of therapeutic

K. Fujinami · A. G. Robson
Institute of Ophthalmology, University College London,
London, UK

A. V. Cideciyan
Center for Hereditary Retinal Degenerations, Scheie Eye
Institute, University of Pennsylvania, Philadelphia, USA

D. A. Thompson
The Tony Kriss Visual Electrophysiology Unit, Clinical
and Academic, Department of Ophthalmology, Sight
and Sound Centre, Great Ormond Street Hospital
for Children NHS Trust, London, UK

D. A. Thompson
Great Ormond Street Institute of Child Health, University
College London, London, UK

C. Matsumoto
Department of Ophthalmology, Kindai University,
Osakasayama, Japan

R. Asaoka
Department of Ophthalmology, Seirei Hamamatsu General
Hospital, Hamamatsu, Shizuoka, Japan

R. Asaoka
Seirei Christopher University, Hamamatsu, Shizuoka,
Japan

R. Asaoka
Nanovision Research Division, Research Institute
of Electronics, Shizuoka University, Shizuoka, Japan

R. Asaoka
The Graduate School for the Creation of New Photonics
Industries, Shizuoka, Japan

C. Johnson
Department of Ophthalmology and Visual Sciences,
University of Iowa, Iowa City, Iowa, USA

C. Johnson
School of Optometry, The Ohio State University,
Columbus, IA, USA

M. W. Dul
Department of Biological and Vision Science, College
of Optometry, State University of New York, New York,
USA

P. H. Artes
Faculty of Health, University of Plymouth, Plymouth, UK

A. G. Robson
Department of Electrophysiology, Moorfields Eye
Hospital, London, UK

interventions, notably in natural history studies of retinal dystrophies, or in clinical trials aimed at arresting retinal degeneration or restoring retinal function [23]. An advantage over more traditional methods is that relatively small decreases or improvements in retinal function may be established, reflecting activity driven by the most sensitive retinal photoreceptors [24]. In contrast, methods such as full-field ERG normally depend on the response of millions of photoreceptors to suprathreshold stimuli, and small changes in retinal function may be undetected, especially in the presence of severe global retinal dysfunction.

Adult patients and older children may undertake FST, likely as potential candidates for clinical trials aimed at monitoring or restoring retinal function, or to monitor safety of therapeutic interventions. Age is an important consideration and may preclude or limit applicability (Sect. "Reduced protocol").

Technical considerations

In order to inform this protocol data were sourced from recent reviews and a systematic review (Online Appendix 2A) (PROSPERO ID 453200) available on <https://www.crd.york.ac.uk/PROSPERO/>, in which papers were graded for quality according to the methodology of a modified Newcastle grading protocol shown in Online Appendix 2B [25]. The papers included in the final review are listed in Table 2 [1, 2, 5, 24, 26–48]. For further details see Online Appendix 3.

The FST is a psychophysical testing method that uses a range of physically well-defined light stimuli to determine visual detection thresholds, i.e. the flash strength corresponding to the stimulus being seen approximately half the time. The spectral and temporal properties of the stimulus, the algorithm used to present sequential stimuli, audible cues, and response choices can contribute strongly to the resulting FST threshold. These and other considerations are outlined below.

Stimulus parameters

Colour

The simplest form of FST uses a white (achromatic) stimulus which provides no information to

differentiate between light sensitivity originating from rods or cones or both. More commonly, two or more spectrally distinct short- and long-wavelength stimuli (blue and red) are used to obtain information regarding different photoreceptor types. Other colours such as green (513 nm) have been used rarely [49]. A white stimulus is produced by using a broadband white light with a colour temperature of 6500 K [29, 31, 36]. Short wavelength (blue) and long wavelength (including red) stimuli can be produced using a range of LEDs [1, 2, 5, 27, 30, 34, 35, 50]. Short wavelength LEDs with wavelengths between 444 and 470 nm and longer wavelength LEDs with wavelengths between 538 and 670 nm have been used.

Temporal stimulus characteristics

Stimulus duration and temporal envelope Stimulus duration (presentation time) is the time between stimulus onset and offset. For clinical visual psychophysical testing, presentation times are usually selected to exceed a critical duration (defined by Bloch's law) of around 100 ms, which defines the stimulus duration beyond which temporal summation no longer influences threshold [51]. In classic (standard automated) perimetry, stimulus durations are usually between 100 ms (Octopus perimeters, Haag-Streit) and 200 ms (Humphrey Field Analyser, Carl Zeiss Meditec), which also is shorter than the typical time taken to execute a saccadic eye movement. With a few exceptions (frequency-doubling technology, flicker perimetry) the temporal envelope of stimuli in perimetry is assumed to be a square-wave [52]. Classic brief duration visual stimuli were produced with a shutter having a sudden onset and offset. Modern FST stimuli are produced with LEDs driven with pulse-width modulation where the current is cycled on and off faster than visual perception. The duration of the LED light stimulus defines both the stimulus strength as well as the duration. Stimulus durations for FSTs reported in the literature range from short variable duration ≤ 4 ms stimuli to longer fixed duration stimuli of 200 ms [2, 29]. Longer stimulus durations allow for greater disease severities to be measured by extending the range to stronger lights, whereas shorter stimuli allow measurement close to normal dark-adapted absolute thresholds by extending the range to dimmer lights.

Interstimulus interval The interstimulus interval (ISI) is the time between the onset of successive stimuli. In conventional perimetry, it is typically around 1400 ms. Many modern thresholding tests use adaptive timing in which the pace of the test is interactively adapted to the speed of the patients' responses [52], providing for a faster and more engaging test. In the context of FST, particularly when dark-adapted, a longer ISI would minimise the influences of sequential stimulus presentations on threshold due to changes in adaptation state. Typically, the FST uses an ISI of less than 2.5 s but as long as 5 s has been reported [30, 39]. However, there is a difference when using a one button versus a two button response system (see Sect. "Response button"). When two buttons are used, the time stops when either button is pressed so the full ISI time window is rarely used. In a system with only one button, the ISI time will be utilised in full when no stimulus is seen and no button pressed, so in this scenario it is important to have the shorter period to avoid unnecessary prolongation of the test.

Response window The response window is the time, relative to stimulus onset, during which responses are accepted as "valid". This window is typically around 800 to 1000 ms for single button testing. Responses that occur implausibly early, e.g. <100 ms after stimulus onset, may be rejected as likely false positive responses [52]. However, it can be difficult to distinguish "anticipatory" false-positive responses from delayed responses to a previous stimulus for single button testing paradigm. The response window is more flexible for two button yes–no testing paradigm.

Clinical protocol

Pupillary dilation

The FST is a measure of maximum sensitivity which depends on the retinal illuminance, the latter being directly proportional to pupil area (Online Appendix 1). FST is thought to require mydriasis (pupil dilation) to help standardise retinal illumination. Mydriasis may reduce inter-subject and inter-session variability, which is of particular relevance in monitoring studies, and has been specified in most of the major studies [1, 2, 5, 26, 27, 29–31, 34–38, 47, 48, 50].

Adaptation

FST can be conducted in both light and/or dark-adapted conditions. Both conditions must be controlled, prior to the start of testing, as the baseline level of retinal light or dark adaptation can influence final thresholds and the test outcome.

Widely different periods of adaptation have been used, ranging from no adaptation to 2 h, [1, 2, 26, 27, 29–32, 34, 36–40, 47, 48, 50] with a median time of approximately 45 min for dark-adapted testing. In disorders characterised by predominant rod dysfunction where dark adaptation may be impaired or delayed, a longer fixed period of dark adaptation may be needed to enable measurement of maximum sensitivity [53]. The second eye tested usually undergoes a longer period of dark adaptation compared to the first eye. Optimising the dark adaptation period for the disease process or considering randomising the order of testing should inform interpretation.

Auditory signal

The use of auditory signals to enhance responses to psychophysical testing has been reported in the perimetry literature [54, 55]. During FST, an auditory prompt signals to the subject when they should respond "yes" (seen) or "no" (not seen) for two button testing paradigm [5, 24, 26, 27, 29, 31, 33, 34, 36, 41–43]. An auditory signal is generally not relevant for one button testing paradigm [32].

Fixation

As the FST is a global response, fixation is not considered important and is rarely reported.

Learning effect

Typically, in psychophysical testing, there is a significant learning effect as individuals adjust to the testing procedure. This has been investigated extensively in perimetry with recommendations to disregard the first three test results in Humphrey field testing [56] and first test for fundus-tracked perimetry [57]. Learning effect has not been formally investigated in FST.

Test–retest variability

Knowledge of test–retest variability is essential to define a significant change following treatment. Roman et al. reported intervisit repeatability of 0.39 log for inherited retinal degeneration patients [1]. Dimopolous et al. reported an intervisit repeatability of 0.27 log and 0.23 log for blue and red stimuli, respectively [5]. Similar test–retest values have been reported for control and patient groups, as well as for all colours. Repeatability should reflect the units of measurement such as log₁₀ units for thresholds measured in log₁₀ units.

Repeat testing

Common practice is to conduct multiple assessments and take the average as the final result. Based on the within-subject standard deviation for white (achromatic stimuli), it has been estimated that four repeats are optimal [1]. A power calculation using a 2-sample, 2-sided *t* test to detect FST changes of 5 dB or more between the means of two sessions using 6 samples each would yield a power of 98% at a significance level of 5% [32]. However, it is important to avoid extended test sessions that may affect results due to subject fatigue. The majority of studies reviewed took an average of 3 measurements [5, 26, 27, 29, 36, 44–46, 50].

Psychometric function and thresholding paradigms

Historically, there have been various methods of performing measurements of visual threshold: (A) the method of adjustment, where the observer varies the strength of a stimulus until it is just detectable, (B) the method of limits, where the observer determines the transition from seeing to non-seeing or vice versa by observing a sequence of ascending or descending light steps until the transition point is reached, (C) the method of constant stimuli in which a series of presentations are provided that are assumed to be above and below the presumed threshold to define a frequency of seeing curve that provides detection performance, and (D) a staircase procedure, in which the stimulus strength is increased and decreased to determine a series of reversals (seeing to non-seeing and vice versa) to identify average detection sensitivity. Each of these procedures has certain advantages and

limitations, and the amount of time to perform the measurements varies considerably [58].

The most comprehensive description of stimulus–response relationships in psychophysical data is the psychometric function. This sigmoidal (s-shaped) function describes visual performance (e.g. % of seen) over a wide range of stimulus strengths. Weak stimuli are associated with a low probability of being seen and strong stimuli with a high response probability. Classically, psychometric functions are measured through the “method of constant stimuli”, in which a stimulus is repeatedly shown in order to establish how often it is seen at each stimulus level tested, according to a pre-determined probability criterion.

For most clinical applications, extreme performance levels (close to 0 or 100%) may not be greatly informative. In these situations, a single point estimate of threshold (or sensitivity) may be a more parsimonious descriptor of performance. The term “threshold” refers to the stimulus strength associated with a particular level of performance, typically 50%, but other values are possible. Sensitivity is the inverse of threshold, but both terms tend to be used interchangeably in the clinical literature.

Thresholds can be derived by statistical modelling (probit, maximum-likelihood, etc.) from frequency-of-seeing type data, or more directly through psychophysical techniques such as staircase, bracketing, or adaptive procedures which depend on the responses during testing [59]. Bayesian thresholding techniques (e.g. Zippy Estimation for Sequential Testing, ZEST) can be constructed to use “prior information” (e.g. the distribution of thresholds in a population), yielding greater efficiency.

The strategies chosen for FST balance speed and accessibility. A majority of studies report a 4–2 staircase thresholding [1, 2, 24, 29, 31, 35, 42, 43], though the number of reversals or steps to final thresholding vary in the literature. Those that report the methodology in the literature may be unaware of commercial proprietary implementation of the thresholding algorithm and may be reporting the previously published method of a 4–2 staircase which may not be the method implemented in proprietary software. Perimetry has seen an evolution of thresholding algorithms and FST methods are also changing. Current methods include an abbreviated method of constant stimuli with pseudo-random expansion to detect the endpoints of the range along the slope of the psychometric

function. Some manufacturers report using an 8–4–2–1 threshold staircase. As with perimetry, careful patient instruction will increase reliability [32].

The most commonly used function for the FST is the two-parameter modified Weibull fit, estimating the stimulus strength associated with 50% probability of detection based on staircase responses (Fig. 1) [5, 26–30, 33, 34, 36–40, 44–48, 50]. The equation of the Weibull function utilised is specified below, though all details of the modifications are not in the public domain and remain proprietary.

$$P(\text{seen}) = 1 - e^{-\left(\frac{\text{strength}}{\text{threshold}}\right)^{\text{slope}}}$$

where P = probability, e = exponential distribution, strength is measured in $\text{cd}\cdot\text{s}\cdot\text{m}^{-2}$, threshold measured in $\text{cd}\cdot\text{s}\cdot\text{m}^{-2}$.

Response button

FST can be performed with a single button response where the patient reports only seen stimuli, or two button response where the patient is typically prompted with an audible cue and reports whether the stimulus is seen or not seen by pressing one of two buttons [32]. In general, single button response offers the advantage of simplicity and patients may be more familiar with the process from previous perimetric testing. Two button testing allows for quantification of both seen and unseen response times, of potential value in quality of data metrics.

Analysis

In visual psychophysics, patient responses reflect not only the underlying sensory capability but also extra-sensory influences such as response criteria, attentiveness, and fatigue [60].

Historically, the patients' response criteria are assessed by interspersing a small number of catch trials in the stimulus sequence. For example, responses to blank trials during which no stimulus was present are interpreted as false positives and may suggest either the patient has a low criterion for responding or experiences photopsias. False positives may also be logged when the patient provides a delayed response to a preceding stimulus that is implausibly early after the next stimulus. False negatives are assessed by the failure to respond to strong stimuli previously

seen [29]. False negatives require an estimate of the threshold and may have limited utility in patients with severe vision loss approaching the limits of the equipment.

The reproducibility of test results can be assessed in terms of test–retest variability via Bland–Altman analyses [61]. A potential limitation arises when variability is markedly non-uniform across the dynamic range of the technique (heteroscedasticity, as in clinical perimetry). In these situations, empirical test–retest intervals can be estimated as an alternative.

Protocol specification

Recommendations are based on a combination of best practice theory and published protocol parameter derived from the systematic review.

Stimulus parameters

Colour

This guideline specifies the use of LEDs with centre wavelengths for the short wavelength (blue) stimulus of between 444 and 470 nm and for long wavelengths (red) between 620 and 670 nm. Once tested, the difference between the red and blue responses should be calculated (see Sect. "Photoreceptor mediation"). White colour temperature should be 6500 K.

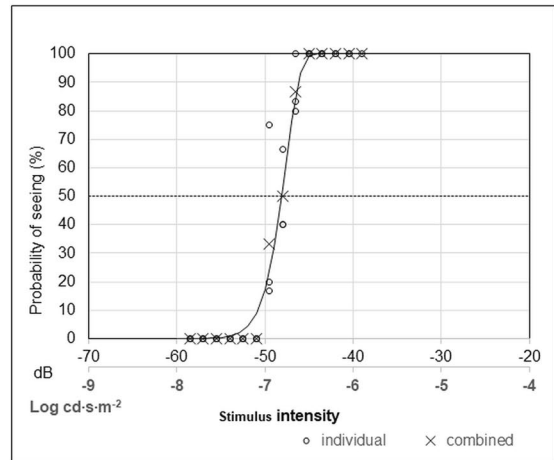
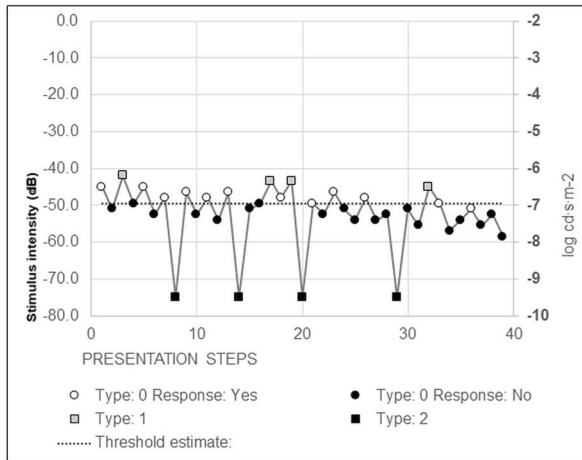
Auditory signal

An auditory cue is recommended for the two button seen–not seen testing paradigm; it is not necessary for single button testing. It is acknowledged that auditory cues may not be appropriate in patients with severe hearing loss and tactile cues may have to be used.

Stimulus duration and interstimulus interval

The stimulus duration must be defined. Literature to date has reported either brief (≤ 4 ms) or longer duration (200 ms) stimuli. Interstimulus interval must be defined. Previous studies have specified an interstimulus interval of approximately 1 s for single button paradigms, or up to 5 s for two button paradigms [29, 47].

A



B

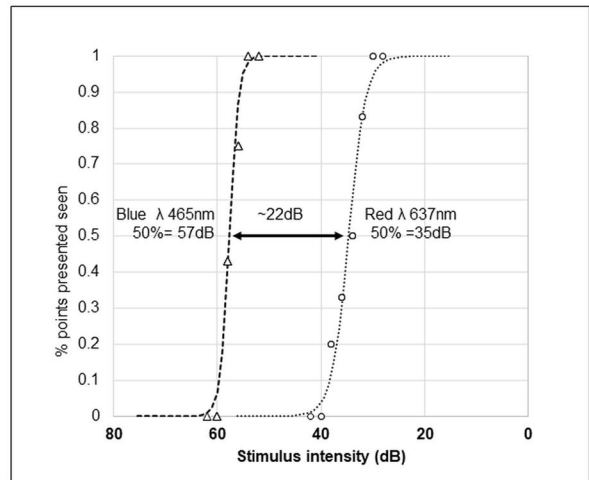
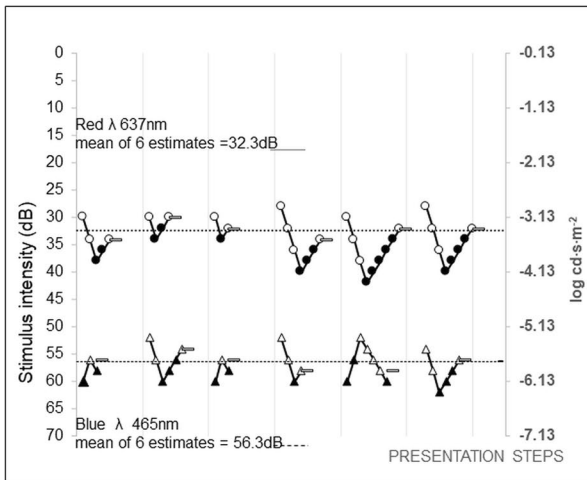


Fig. 1 Examples of seen–unseen two button responses with corresponding modified Weibull function fitted to the data. (A) Top panel shows data collection following 45-min dark adaptation to a white (6500 k) stimulus with a variable duration of less than 4 ms presented in a ganzfeld (reference 0 dB = 0.01 cd·s·m⁻²). A proprietary presentation order of stimuli is demonstrated on the left panel, in this instance with 8 catch trials. A two button response indicates whether the stimulus is seen or unseen within a 5 s response window after an audible cue. The right panel shows probability or percentage of seen stimuli plotted against stimulus strength (“intensity”) and fitted by a modified Weibull function to the data combined from individual trials, which provides a 50% seeing mean estimate of 48 dB [−6.8 log cd·s·m⁻²]. (B) The bottom panel displays six consecutive measurements using 4–2 stair-

case method with 2 reversals in each case. Blue (465 nm) and red (637 nm) flash stimuli with a fixed duration of 200 ms were presented following 45-min dark adaptation in a ganzfeld stimulator (reference 0 dB = 3.7 cd·m⁻²). A single button press indicates seen. Plotted stimuli seen are indicated by open symbols, not seen by filled symbols. Triangles and circles for blue and red stimuli, respectively. The stimulus strength of the last seen is taken as the threshold, marked with a horizontal line. This process is repeated 6 times for each stimulus and an average of thresholds and standard deviation taken. The corresponding plotted threshold and modified Weibull functions are shown in the panel on the right, to provide an alternative, but similar, estimates of the 50% seen thresholds of 57 dB blue and 35 dB red, with a blue–red difference of 22 dB

Break periods should be provided between bouts of testing to allow re-adaptation to the dark and to reduce fatigue. The length of these should be defined, ideally > 3 min.

Clinical protocol

Dilation

Pupil dilation is a requirement of this guideline for clinical trials and research. Pupil diameters should be measured and recorded. If mydriasis is contraindicated, this must be acknowledged as a departure from the guideline and the likely influence on thresholds and interpretation of results considered.

Adaptation

For light-adapted testing, a minimum light adaptation period of 5 min is recommended. Prior to dark-adapted testing, a period of 45-min dark adaptation is recommended, subject to consideration of the patient's diagnosis. This can be assessed by checking the dark adaptation characteristics in a sample of disease patients prior to finalising the protocol. In retinal disorders characterised by severely delayed dark adaptation, a longer fixed period of dark adaptation may be used, if necessary to obtain detectable or robust responses, although care must be taken to ensure consistency between patients and serial assessments.

Response button

Whether a one button or two button box is used and the manner in which it is used should be specified, such as alternative choice with an auditory cue.

Patching

Adequate patching of the contralateral eye to prevent light stimulus "leakage" is essential and may require use of additional patching to provide blackout conditions. It is highlighted that the sensitivity of one eye may be orders of magnitude higher than the fellow, e.g. in unocular gene-based treatment trials. The order of eye testing should be reported as the second tested eye may have longer dark adaptation.

Psychometric function and thresholding paradigms

Thresholding algorithm

The algorithm for sequential presentation of stimuli should be specified [1, 2, 24, 29, 31, 35, 42, 43]. The observer or patient should be instructed clearly to optimise compliance and reliability of measurements. FST is a psychophysical test and patient instruction has a profound impact on the reliability of testing and final threshold obtained. On this basis, an example of good practice instructions is provided in Online Appendix 4.

Starting stimulus

Under dark-adapted testing, longer wavelength (red) stimuli should be presented first followed by shorter wavelength (blue) stimuli, to minimise the disruption of the dark-adapted state of the retina.

It can be helpful to commence testing closer to the expected final threshold where possible, to reduce test time.

Threshold

Threshold should be defined as the stimulus strength associated with 50% probability of detection, based on the psychometric function, e.g. two-parameter modified Weibull fit.

Analysis, interpretation, and reporting

Learning effect

Observers should perform several trial runs for training, e.g. one to two minutes, immediately prior to testing. More are recommended if needed to optimise patient comfort and compliance.

Catch trials

False positive catch trials should be included in the test in order to provide some indication of performance reliability during testing. False negative catch trials if recorded should be interpreted with caution as they are a suprathreshold presentation and have the potential to disrupt the dark adaptive state.

Test–retest variability

Data analysis requires an appreciation of test–retest variability and this should be examined for the system and stimuli being used, and for the population being studied.

Repeat testing

If the first three tests are consistent, then three are sufficient to comply with this guideline. In cases of high intra-subject variability, attempts should be made to encourage better compliance and further repeats obtained to optimise consistency. This applies separately to each test condition used.

Units of measurement

Threshold results should be reported in units of $\log \text{cd}\cdot\text{s}\cdot\text{m}^{-2}$ for brief stimuli, or in units of $\log \text{cd}\cdot\text{m}^{-2}$ with longer duration stimuli together with the stimulus duration. Considering the use of chromatic stimuli under dark- and light-adapted conditions, units must be specified in photopic or scotopic units (phot-cd or scot-cd) as appropriate to the application. Decibels units ($\text{dB} = 10\log_{10}$) should not be used, as although widely employed in psychophysical techniques such as perimetry, it is a relative scale dependent on the 0 dB point (reference stimulus strength).

Photoreceptor mediation

In two-colour chromatic FST testing, the difference in thresholds between short- and long-wavelength tests can provide an estimate of the photoreceptors contributing to each response. The precise difference will partly depend on the radiometric properties of the LED stimuli, and the equipment used must therefore be specified [2].

Reference ranges

Establishing reference (“normative”) values involves recruiting and testing sufficient reference subjects per clinically relevant partition, and establishing laboratory-specific reference limits is generally considered the optimal process. If external or published reference data (e.g. [62]) are to be used they must be verified as appropriate for the local methods and equipment,

with an understanding of possible limitations and how reference limits were defined.

Quality control

The quality and consistency of measurements should be monitored and reported. This could be in the form of monitoring the quality of the responses as they were collected and repeating if required and/or excluding unreliable or low-quality tests, although if necessary this should be clearly acknowledged. The consistency of data and shape of the psychometric function fitted to the measurements should be examined to assess the reliability of the responses (Online Appendix 5). If a disproportionate number of points are located around the 0% and 100% locations compared to intermediate values along the probability curve, this may indicate a skew in the response characteristics and impair curve fitting.

Reporting recommendations checklist

In order to facilitate comparison between reports, the following minimum information should be specified in all FST reports:

1. Adherence to this guidance should be stated, and any departures acknowledged and justified.
2. Equipment used and full details of test paradigm employed.
3. For proprietary software, if full details are not available, it is essential to specify the version number, settings and protocol.
4. Stimulus duration and interstimulus interval.
5. Stimulus colour including the peak wavelength of LEDs used.
6. Pupil sizes after mydriasis.
7. Duration of light and dark adaptation including periods of re-adaptation.
8. The quality assessments employed.
9. The number of tests completed for each stimulus condition.
10. The eye tested and order of eyes tested.
11. The response method (e.g. use of a one or two button box and test paradigm) and thresholding algorithm.
12. Thresholds provided in $\log \text{cd}\cdot\text{s}\cdot\text{m}^{-2}$ for brief flash durations under 4 ms, and in $\log \text{cd}\cdot\text{m}^{-2}$ for longer durations.

Reduced protocol

When comprehensive FST is not possible, such as for young children, those with special needs, and adults unable to comply with routine testing, a reduced protocol may be considered. Modifications may have a negative impact on reliability and sensitivity, but there remains potential to yield meaningful and clinically useful results, providing the core principles of FST are retained.

It is well recognised that paediatric psychometric testing of vision can be challenging, depending on the age, capability and compliance of the child. Maturity and compliance can vary significantly between children of the same age and across ages requiring examiner's judgement [63]. It has been reported that children older than 8 years are able to reliably perform perimetry [64]. Published FST data suggest children over 6 years without physical or neurodevelopmental impairment may successfully complete the FST. Those under 6 years, or older children/adults unable to understand the requirements of the test, or unable to physically comply with a button press may need support or adaptations.

Any test modification may impact the reliability and sensitivity of a test and adaptations must be documented carefully so that the test protocol can be replicated for monitoring. Compliance must be detailed and recorded to inform data interpretation. At this time, there are no evidence-based age-related modifications for children and this is an area that requires further research. This guidance is based on guidance from other psychophysical testing.

Shortened protocols will usually involve changing at least one of several parameters. Ways of making the test into a game, involving the child's carer or an older sibling, and allowing breaks for drinks and snacks between tests may be helpful. The period of dark adaptation may be shortened, or testing performed without mydriasis. There may be fewer types of stimuli per test session such as prioritising white over colour stimuli, depending upon the clinical priority, with shorter but more frequent test sessions. Simultaneous rather than sequential testing of both eyes and simpler response protocols may be considered, e.g. using one button rather than two or allowing verbal or tactile responses that a carer reports back. This may require a longer ISI to provide time for the feedback and the carer must be

counselled not to bias responses. Balancing quality control measures, such as decreasing the number of catch trials, with test duration is also important.

Acknowledgements We would like to thank Linda Shi for her help with the systematic review and the ISCEV membership for feedback, following presentations of draft protocols at the ISCEV annual symposia in 2022 and 2023. We are grateful to the manufacturers for the technical information provided to complete this document, to Alejandro Roman for providing the data for Fig 1B and to Shazeea Ishmael for administrative support.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Roman AJ, Schwartz SB, Aleman TS et al (2005) Quantifying rod photoreceptor-mediated vision in retinal degenerations: dark-adapted thresholds as outcome measures. *Exp Eye Res* 80:259–272. <https://doi.org/10.1016/j.exer.2004.09.008>
2. Roman AJ, Cideciyan AV, Aleman TS, Jacobson SG (2007) Full-field stimulus testing (FST) to quantify visual perception in severely blind candidates for treatment trials. *Physiol Meas* 28:N51–N56. <https://doi.org/10.1088/0967-3334/28/8/N02>
3. Klein M, Bennett LD, Kiser K et al (2016) A comparison of the medmont dark adapted chromatic perimeter (DAC) with the full-field stimulus threshold (FST) in subjects with retinitis pigmentosa (RP). *ARVO annual meeting abstract*. *Invest Ophthalmol Vis Sci* 57(12):631
4. Simunovic MP, Grigg JR, Mahroo OA (2022) Vision at the limits: absolute threshold, visual function, and outcomes in clinical trials. *Surv Ophthalmol* 67:1270–1286
5. Dimopoulos IS, Freund PR, Knowles JA, MacDonald IM (2018) The natural history of full-field stimulus threshold decline in choroideremia. *Retina* 38:1731–1742. <https://doi.org/10.1097/IAE.0000000000001764>
6. Jolly JK, Bridge H, MacLaren RE (2019) Outcome measures used in ocular gene therapy trials: a scoping review of current practice. *Front Pharmacol* 10:1076. <https://doi.org/10.3389/fphar.2019.01076>

7. Hamilton R, Bach M, Heinrich SP et al (2021) ISCEV extended protocol for VEP methods of estimation of visual acuity. *Doc Ophthalmol* 142:17–24. <https://doi.org/10.1007/s10633-020-09780-1>
8. McCulloch D, Bach M, Brigell M et al (2023) ISCEV Guidelines to instrument for calibration and verification of stimuli and recording instruments. *Documenta Ophthalmologica* 146:199–210
9. Constable PA, Bach M, Frishman LJ et al (2017) ISCEV Standard for clinical electro-oculography (2017 update). *Doc Ophthalmol* 134:1–9
10. Brigell M, Jeffrey BG, Mahroo OA, Tzekov R (2020) ISCEV extended protocol for derivation and analysis of the strong flash rod-isolated ERG a-wave. *Doc Ophthalmol* 140:5–12
11. Perlman I, Kondo M, Chelva E et al (2020) ISCEV extended protocol for the S-cone ERG. *Doc Ophthalmol* 140:95–101
12. McCulloch DL, Kondo M, Hamilton R et al (2019) ISCEV extended protocol for the stimulus–response series for light-adapted full-field ERG. *Doc Ophthalmol* 138:205–215
13. Johnson MA, Jeffrey BG, Messias AMV, Robson AG (2019) ISCEV extended protocol for the stimulus–response series for the dark-adapted full-field ERG b-wave. *Doc Ophthalmol* 138:217–227
14. Sustar M, Holder GE, Kremers J et al (2018) ISCEV extended protocol for the photopic On–Off ERG. *Doc Ophthalmol* 136:199–206
15. Thompson DA, Fujinami K, Perlman I et al (2018) ISCEV extended protocol for the dark-adapted red flash ERG. *Doc Ophthalmol* 136:191–197
16. Frishman L, Sustar M, Kremers J et al (2018) ISCEV extended protocol for the photopic negative response (PhNR) of the full-field electroretinogram. *Doc Ophthalmol* 136:207–211
17. Bach M, Brigell MG, Hawlina M et al (2013) ISCEV standard for clinical pattern electroretinography (PERG): 2012 update. *Doc Ophthalmol* 126:1–7
18. Hoffmann MB, Bach M, Kondo M et al (2021) ISCEV standard for clinical multifocal electroretinography (mfERG)(2021 update). *Doc Ophthalmol* 142:5–16
19. Odom JV, Bach M, Brigell M et al (2016) ISCEV standard for clinical visual evoked potentials:(2016 update). *Doc Ophthalmol* 133:1–9
20. Robson AG, Nilsson J, Li S et al (2018) ISCEV guide to visual electrodiagnostic procedures. *Doc Ophthalmol* 136:1–26
21. Robson AG, Frishman LJ, Grigg J et al (2022) ISCEV Standard for full-field clinical electroretinography (2022 update). *Doc Ophthalmol* 144:165–177
22. Sample PA, Dannheim F, Artes PH et al (2011) Imaging and perimetry society standards and guidelines. *Optom Vis Sci* 88:4–7. <https://doi.org/10.1097/OPX.0b013e3181fc3735>
23. Ayton LN, Rizzo JF, Bailey IL et al (2020) Harmonization of outcomes and vision endpoints in vision restoration trials: Recommendations from the international HOVER taskforce. *Transl Vis Sci Technol* 9:25. <https://doi.org/10.1167/tvst.9.8.25>
24. Cideciyan AV, Jacobson SG, Ho AC et al (2021) Durable vision improvement after a single treatment with anti-sense oligonucleotide sefofarsen: a case report. *Nat Med* 27:785–789. <https://doi.org/10.1038/s41591-021-01297-7>
25. Wells GA, Shea B, O’Connell D, et al The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 5 Apr 2023
26. Collison FT, Fishman GA, McAnany JJ et al (2014) Psychophysical measurement of rod and cone thresholds in stargardt disease with full-field stimuli. *Retina* 34:1888–1895. <https://doi.org/10.1097/IAE.000000000000144>
27. Collison FT, Park JC, Fishman GA et al (2015) Full-field pupillary light responses, luminance thresholds, and light discomfort thresholds in CEP290 leber congenital amaurosis patients. *Invest Ophthalmol Vis Sci* 56:7130–7136. <https://doi.org/10.1167/iovs.15-17467>
28. Jacobson SG, Cideciyan AV, Aleman TS et al (2011) Human retinal disease from AIPL1 gene mutations: foveal cone loss with minimal macular photoreceptors and rod function remaining. *Invest Ophthalmol Vis Sci* 52:70–79. <https://doi.org/10.1167/iovs.10-6127>
29. Klein M, Birch DG (2009) Psychophysical assessment of low visual function in patients with retinal degenerative diseases (RDDs) with the Diagnosys full-field stimulus threshold (D-FST). *Doc Ophthalmol* 119:217–224. <https://doi.org/10.1007/s10633-009-9204-7>
30. McAnany JJ, Park JC, Fishman GA, Collison FT (2020) Full-field electroretinography, pupillometry, and luminance thresholds in X-linked retinoschisis. *Invest Ophthalmol Vis Sci* 61:53. <https://doi.org/10.1167/IOVS.61.6.53>
31. Messias K, Jägle H, Saran R et al (2013) Psychophysically determined full-field stimulus thresholds (FST) in retinitis pigmentosa: relationships with electroretinography and visual field outcomes. *Doc Ophthalmol* 127:123–129. <https://doi.org/10.1007/s10633-013-9393-y>
32. Roman AJ, Cideciyan AV, Wu V et al (2022) Full-field stimulus testing: Role in the clinic and as an outcome measure in clinical trials of severe childhood retinal disease. *Prog Retin Eye Res* 87:101000. <https://doi.org/10.1016/j.preteyeres.2021.101000>
33. Stingl KT, Kuehlewein L, Weisschuh N et al (2019) Chromatic full-field stimulus threshold and pupillography as functional markers for late-stage, early-onset retinitis pigmentosa caused by *crbl1* mutations. *Transl Vis Sci Technol* 8:45. <https://doi.org/10.1167/tvst.8.6.45>
34. Testa F, Melillo P, Della CM et al (2021) Voretigene neparvovec gene therapy in clinical practice: treatment of the first two Italian pediatric patients. *Transl Vis Sci Technol* 10:45. <https://doi.org/10.1167/tvst.10.10.11>
35. William A, Blak MD, Eshref A et al (2022) Chromatic full-field stimulus thresholds in patients with treatment-naive age-related macular degeneration. *Clin Ophthalmol* 16:223–229. <https://doi.org/10.2147/OPHTH.S346291>
36. Zabek O, Calzetti G, Prétot D et al (2022) Full-field sensitivity threshold and the relation to the oxygen metabolic retinal function in retinitis pigmentosa. *Graefes Arch Clin Exp Ophthalmol* 260:2517–2527. <https://doi.org/10.1007/s00417-022-05638-1>

37. Humayun MS, Dorn JD, Da Cruz L et al (2012) Interim results from the international trial of second sight's visual prosthesis. *Ophthalmology* 119:779–788. <https://doi.org/10.1016/j.ophtha.2011.09.028>
38. Ghazi NG, Abboud EB, Nowilaty SR et al (2016) Treatment of retinitis pigmentosa due to MERTK mutations by ocular subretinal injection of adeno-associated virus gene vector: results of a phase I trial. *Hum Genet* 135:327–343. <https://doi.org/10.1007/s00439-016-1637-y>
39. Bennett J, Wellman J, Marshall KA et al (2016) Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. *Lancet* 388:661–672. [https://doi.org/10.1016/S0140-6736\(16\)30371-3](https://doi.org/10.1016/S0140-6736(16)30371-3)
40. Maguire AM, Russell S, Wellman JA et al (2019) Efficacy, safety, and durability of voretigene neparvovec-rzyl in RPE65 mutation-associated inherited retinal dystrophy: results of phase 1 and 3 trials. *Ophthalmology* 126:1273–1285. <https://doi.org/10.1016/j.ophtha.2019.06.017>
41. Cideciyan AV, Jacobson SG, Drack AV et al (2019) Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. *Nat Med* 25:225–228. <https://doi.org/10.1038/s41591-018-0295-0>
42. Cideciyan AV, Jacobson SG, Ho AC et al (2022) Restoration of cone sensitivity to individuals with congenital photoreceptor blindness within the phase 1/2 sepfarsen trial. *Ophthalmol Sci* 2:100133. <https://doi.org/10.1016/j.xops.2022.100133>
43. Russell SR, Drack AV, Cideciyan AV et al (2022) Intravitreal antisense oligonucleotide sepfarsen in Leber congenital amaurosis type 10: a phase 1b/2 trial. *Nat Med* 28:1014–1021. <https://doi.org/10.1038/s41591-022-01755-w>
44. Birch DG, Cheng P, Duncan JL et al (2020) The rush2a study: best-corrected visual acuity, full-field electroretinography amplitudes, and full-field stimulus thresholds at baseline. *Transl Vis Sci Technol* 9:1–12. <https://doi.org/10.1167/tvst.9.11.9>
45. Birch DG, Samarakoon L, Melia M et al (2022) The RUSH2A study: dark-adapted visual fields in patients with retinal degeneration associated with biallelic variants in the USH2A gene. *Invest Ophthalmol Vis Sci* 63:17. <https://doi.org/10.1167/iovs.63.3.17>
46. Hufnagel RB, Liang W, Duncan JL et al (2022) Tissue-specific genotype–phenotype correlations among USH2A-related disorders in the RUSH2A study. *Hum Mutat* 43:613–624. <https://doi.org/10.1002/humu.24365>
47. Ahuja AK, Yeoh J, Dorn JD et al (2013) Factors affecting perceptual threshold in argus II retinal prosthesis subjects. *Transl Vis Sci Technol* 2:1. <https://doi.org/10.1167/tvst.2.4.1>
48. Ahuja AK, Behrend MR (2013) The Argus™ II retinal prosthesis: Factors affecting patient selection for implantation. *Prog Retin Eye Res* 36:1–23
49. Yamamoto M, Matsuyama T, Maeda T et al (2023) Detailed evaluation of chromatic pupillometry and full-field stimulus testing to assess ultralow vision in retinitis pigmentosa. *Ophthalmol Sci* 3:100328. <https://doi.org/10.1016/j.xops.2023.100328>
50. Talib M, van Schooneveld MJ, Wijnholds J et al (2021) Defining inclusion criteria and endpoints for clinical trials: a prospective cross-sectional study in CRB1-associated retinal dystrophies. *Acta Ophthalmol* 99:e402–e414. <https://doi.org/10.1111/aos.14597>
51. Baumgardt E (1972) Threshold quantal problems. In: Alpern M, Aulhorn E, Barlow HB et al (eds) *Visual psychophysics*. Springer, Berlin, pp 29–55
52. Bengtsson B, Olsson J, Heijl A, Rootzen H (1997) A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand* 75:368–375. <https://doi.org/10.1111/j.1600-0420.1997.tb00392.x>
53. Cideciyan AV, Aleman TS, Boye SL et al (2008) Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc Natl Acad Sci* 105:15112–15117. <https://doi.org/10.1073/pnas.0807027105>
54. Doyle MC, Snowden RJ (2001) Identification of visual stimuli is improved by accompanying auditory stimuli: the role of eye movements and sound location. *Perception* 30:795–810. <https://doi.org/10.1068/p3126>
55. Buller J, Rodriguez J, Miller J (2013) The effect of automated auditory cues on humphrey visual field performance. Undergraduate Dissertation, Northeastern State University, Oklahoma
56. Gillespie BW, Musch DC, Guire KE et al (2003) The collaborative initial glaucoma treatment study: Baseline visual field and test-retest variability. *Invest Ophthalmol Vis Sci* 44:2613–2620. <https://doi.org/10.1167/iovs.02-0543>
57. Buckley TMW, Jolly JK, Menghini M et al (2020) Test-retest repeatability of microperimetry in patients with retinitis pigmentosa caused by mutations in RPGR. *Clin Exp Ophthalmol* 48:714–715. <https://doi.org/10.1111/ceo.13753>
58. Hu R, Racette L, Chen KS, Johnson CA (2020) Functional assessment of glaucoma: uncovering progression. *Surv Ophthalmol* 65:639–661. <https://doi.org/10.1016/j.survophthal.2020.04.004>
59. Treutwein B (1995) Adaptive psychophysical procedures. *Vis Res* 35:2503–2522. [https://doi.org/10.1016/0042-6989\(95\)00016-X](https://doi.org/10.1016/0042-6989(95)00016-X)
60. Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. John Wiley, Oxford
61. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160. <https://doi.org/10.1177/096228029900800204>
62. Shi LF, Hall AJ, Thompson DA (2023) Full-field stimulus threshold testing: a scoping review of current practice. *Eye*. <https://doi.org/10.1038/s41433-023-02636-3>
63. Patel DE, Cumberland PM, Walters BC et al (2015) Study of optimal perimetric testing in children (OPTIC): Feasibility, reliability and repeatability of perimetry in children. *PLoS ONE* 10:e0130895
64. Blumenthal E, Haddad A, Horani A, Anteby I (2004) The reliability of frequency-doubling perimetry in young children. *Ophthalmology* 111:435–439. <https://doi.org/10.1016/j.ophtha.2003.06.018>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.