



Instructor's Guide to ICAR Cognitive Reasoning

Laboratory implementation, theoretical foundations,
experimental evidence, and variants

Behavioral and Experimental Economics Research Lab
Department of Economics



Revised edition | October 2026

<https://beer-lab.org/>

How to use this guide

This guide supports preparation, administration and discussion of the BEER Lab ICAR cognitive reasoning activity. It is intended for instructors in economics, research methods and behavioral science. Its purpose is to explain how performance is observed and summarized under explicit scoring and missing-data rules.

There are three reading routes: the operational chapters support rehearsal; the foundations explain interpretation; the worked cases develop denominator checks and limits of inference. The original 16-item form and extended 60-item form are administered separately, once per person.

These guides do not reproduce the bank or its answer keys. Original materials require the scientific license accepted for this teaching purpose. The implementation keeps them private and documents administration adaptations and corrections to the distributed Spanish workbook.

Author

This guide was prepared by Stanislaw Maldonado, research professor in the Department of Economics at Tecnológico de Monterrey, Monterrey Campus. For questions or suggestions about its content, please contact the author at stanislao.maldonado@tec.mx.

For the instructor

Rehearse with synthetic data, confirm language and edition, and close the access-code panel before projection. Preserve form, language and key. Distinguish correct, incorrect, do-not-know, omitted and pending. State N and missing-data rules before interpreting differences.

Contents

How to use this guide	i
1 The activity and its administration	1
1.1 What is administered and what the class should teach	1
1.2 Institutional navigation and room selection	2
1.3 Local rehearsal and participant links	2
1.4 Instructor credentials and activity access code	4
1.5 Instructions and comprehension check	4
1.6 One problem per screen and response recording	4
1.7 Figures, translation and practical accessibility	4
1.8 Completion and individual feedback	7
1.9 The private report before disclosure	7
1.10 Reading the revealed report	7
1.11 Downloading and checking data	8
1.12 Lesson planning and pedagogical closing	11
1.13 Materials and continuity	11
2 Foundations of measurement and scoring	12
2.1 From construct to observable record	12
2.2 Four problem types	12
2.3 Content coverage and form composition	13
2.4 Binary scoring and distinct response states	13
2.5 The total and components have separate N	13
2.6 Center and dispersion	14
2.7 Intervals, percentages and item review	14
2.8 Timing and administration conditions	15
2.9 What the original sources establish	15
3 Evidence and limits of interpretation	16
3.1 Reliability, validity and different uses	16
3.2 Measurement error and uncertainty	16
3.3 Translation, equivalence and declared corrections	16
3.4 Effort, incentives and prior exposure	17
3.5 Connections to behavioral economics	17
3.6 Comparing forms and conditions	18
3.7 A learning rubric	18
4 Worked cases for interpretation practice	19
5 References and source materials	25

List of Figures

1.1	Four rooms separate language and edition.	2
1.2	The basic Spanish room offers only its corresponding configuration.	3
1.3	Demo shows four configurations isolated to this family.	3
1.4	Neutral welcome and entry using the session access code.	5
1.5	Instructions, permitted resources and problem count.	6
1.6	Procedural feedback explains correction while preserving the choice.	6
1.7	Closing confirms recording while feedback remains hidden.	7
1.8	Extended private feedback requires completion and current disclosure.	8
1.9	The instructor chooses when results become visible.	9
1.10	Each measure retains its own maximum and observation count.	10
1.11	Contiguous total-score intervals have frequencies summing to N.	10

The activity and its administration

Purpose	Teach cognitive performance measurement, scoring and limits of inference.
Experimental unit	Person; individual activity without matching or waiting.
Basic edition	Original 2014 ICAR-16: four items in each of four types.
Extended edition	Original ICAR-60: 9 series, 16 verbal, 11 matrix and 24 rotation items. Each person receives one form.
Decision	Choose one alternative, or explicitly omit.
Outcome	One point per correct answer; zero for submitted wrong, do-not-know or omitted answers. Pending items prevent a complete score. No payments or IQ conversion.
Topics	cognitive reasoning; cognitive ability; verbal reasoning; spatial reasoning; performance measurement
JEL	A22, C83, C91, D91, I21
Participants	Any positive count; one person per interface.
Suggested duration	Allow 15-20 minutes for basic and 35-50 for extended; pilot locally. No automatic limit.

<https://beer-lab.org/>

<https://teaching.beer-lab.org/>

https://icar.beer-lab.org/room/icar_en

<http://localhost:9173/demo>

1.1 What is administered and what the class should teach

The activity presents problems with one correct alternative and preserves the choice made. The basic edition is the original 16-item subset; the extended edition is the original 60-item form. Do not administer the short form and then treat the same person's full-form result as an independent observation. They share problems, and a second exposure introduces familiarity. Choose the edition according to available time and the curricular objective: an introduction to measurement or a longer examination of composition and precision.

The observational unit is a person. There are no partners, social roles, synchronization, endowments, payments or demographic questions. Each participant progresses without waiting for others. The instructor can use results to explain scoring, distributions and selection of observations. The

activity by itself does not generate an individual diagnosis or an academic grade. Announce that broad purpose before beginning without suggesting solution strategies, expected performance or desirable answers.

1.2 Institutional navigation and room selection

Institutional access begins at <https://beer-lab.org/> and continues through its teaching section to <https://teaching.beer-lab.org/>. Once published, the family card identifies its title, topics and editions. Proposed rooms are `icar_es` and `icar_en` for the basic editions, and `icar_extendido` and `icar_extended` for the extended editions. A proposed public address does not establish that production deployment is already complete. Check the host and room before distributing instructions.

This family's instance should contain four configurations in Demo and four rooms in Rooms. Each room selects one configuration with the corresponding language and edition. A basic room must not default to an extended configuration. If unrelated games appear, a selector is empty or the language is incorrect, stop distributing links and check the instance. This small operational check prevents substantial confusion between cohorts, forms and databases.

Rooms

Current rooms:

ICAR-16 · Español · básica
ICAR-16 · English · basic
ICAR-60 · Español · extendida
ICAR-60 · English · extended

Figure 1.1: Four rooms separate language and edition.

1.3 Local rehearsal and participant links

The delivery preview uses <http://localhost:9173/demo> and works only on the computer running that instance. The intended host is <https://icar.beer-lab.org/>; access from other devices requires the corresponding deployment. A room address is a shared entry route, an individual link identifies one position within a session, and a report is a private instructor page. These addresses are not interchangeable. Giving two people the same individual link mixes their record even if storage itself works correctly.

Prepare a rehearsal with demonstration responses separately from the classroom session. Check language, problem count, figures and closing from a participant position. Then create a new session for the expected attendance and record its identifier. Any positive number of participants is admissible. Using three positions for a rehearsal helps check a complete record, a partial record and an untouched position; it does not establish an experimental group size.

Room: ICAR-16 · Español · básica

Create a new session

Session Config

ICAR-16 · Español · básica ▾

Number of participants

Decisiones individuales: se admite cualquier número positivo de participantes.

Create

[Configure session](#) ⚙️

App sequence

icar_assessment (60 rounds)

Separate original ICAR-16/60 forms; private material input, descriptive counts and no IQ conversion.

0 participants present

Start links

These URLs will stay constant for new sessions, even if the database is recreated.

Room-wide URL

Here is the room-wide URL anyone can use. Don't use this link when testing in multiple tabs on the same browser, because all tabs will be assigned to the same participant, using a cookie.

http://localhost:9173/room/icar_es

Figure 1.2: The basic Spanish room offers only its corresponding configuration.

Demo

ICAR-16 · Español · básica

ICAR-16 · English · basic

ICAR-60 · Español · extendida

ICAR-60 · English · extended

**Razonamiento
cognitivo ICAR
/ ICAR
Cognitive
Reasoning —
BEER Lab**

Seleccione una variante
de este juego. Select a
version of this game.

Figure 1.3: Demo shows four configurations isolated to this family.

1.4 Instructor credentials and activity access code

Use the institutional administrator credentials and keep them outside materials distributed to students. Each session generates an activity access code, available inside an expandable report panel. This code differs from the administrator password and technical session identifier. Share it with the intended cohort and close the panel before projection. The code restricts operational entry; it does not make individual links appropriate for public circulation.

The welcome requires the code and acceptance of general laboratory rules. An incorrect code produces an explanation and allows correction. After successful entry, the application clears the field and records access. Someone who has not entered remains pending rather than having zero correct answers. Preserve this distinction when checking progress or downloading responses. The activity code must not appear in response exports.

1.5 Instructions and comprehension check

Instructions require independent work without notes, calculators, internet searches or other aids. Alphabetic series use the 26 letters A through Z. The application does not impose a countdown or convert elapsed seconds into points. It also explains explicit omission and the absence of payments or academic grading based on answers. Do not announce solutions, hidden rules or comparisons of ability between classmates.

The comprehension check verifies procedure rather than ability to solve an item. An incorrect procedural choice about external aids receives a specific explanation and can be corrected. Resolve navigation questions without training participants on bank problems. Comprehension errors are not added to cognitive scores. Their count helps review operational clarity and cannot substitute for a validated comprehension measure.

1.6 One problem per screen and response recording

Each screen presents a prompt, a figure when needed and the problem's alternatives. Substantive options preserve their source mapping. Matrix problems contain six graphical alternatives and the additional source options; rotations also contain none-of-the-cubes and do-not-know choices. Do not assume that a position means the same thing across types: none of the rotations is D in rotation items. Scoring uses each item's key rather than a common positional rule.

The interface adds explicit omission separately from do-not-know. Confirmation records the response and advances. This teaching adaptation groups series, verbal problems, matrices and rotations, with ascending original identifiers inside each block. The fixed order supports classroom follow-up and is documented as an adaptation; the original sample recommends randomization. Observed position patterns here must not be described as results of a randomized order experiment.

1.7 Figures, translation and practical accessibility

Graphical stimuli remain private original files checked by hashes. In Spanish, native HTML labels replace the displayed English D and H text without redrawing cubes or changing their symbols. Selectable alternatives are also translated. Check presentation on the devices students actually use. A figure can fit a screen while remaining too small to inspect relations between faces.

Allow browser enlargement when needed and check that alternatives and confirmation controls



Individual activity

The Behavioral and Experimental Economics Research Lab in the Department of Economics at Tecnológico de Monterrey is a place to learn about economic decision making through controlled activities.

Participation rules

- Work individually and privately.
- Remain silent and do not discuss your responses with others.
- Follow the instructions on each screen.
- Do not refresh or close the page while responding.
- If you need help, raise your hand or contact staff privately.

Responses are private. The instructor may show aggregate results after choosing to reveal them. Individual profiles will not be projected.

Access code provided by the instructor

☐ I have read the rules and wish to start.

Start

Figure 1.4: Neutral welcome and entry using the session access code.

BEER Lab

Instructions

Work independently without notes, calculators, the internet or other aids. The activity has no automatic time limit. Letter series use the 26-letter A-Z alphabet.

Select one option per problem. Work without a calculator, online searches or external assistance. You may indicate that you prefer not to answer. Each submission confirms your choice and advances to the next problem. There is no automatic time limit or payment; answers do not determine an academic grade.

This is an individual activity. There are no partners, endowments or payments linked to responses. You may indicate that you prefer not to answer an item.

Responses are private. The instructor may show aggregate results after choosing to reveal them. Individual profiles will not be projected.

60 problems.

Continue

Figure 1.5: Instructions, permitted resources and problem count.

BEER Lab

Before you start

How should you solve the problems?

- ☐ Individually, without external tools or assistance
- ☒ By searching online or discussing answers with other people

Your answer: By searching online or discussing answers with other people. The correct answer is Individually, without external tools or assistance. Solve the problems individually, without a calculator, online searches or external assistance. You may use explicit omission if you prefer not to answer.

Continue

Figure 1.6: Procedural feedback explains correction while preserving the choice.

remain accessible. An alternative description identifies the figure but does not turn a spatial visual task into an equivalent assessment for someone unable to examine it. Access accommodations require planning and documentation. Do not attribute difficulties caused by displays, language or connectivity to a stable personal characteristic. When conditions prevent adequate participation, record the incident and preserve incomplete problem states.

1.8 Completion and individual feedback

The basic form closes after its 16 active problems. The application uses 60 technical periods to share logic with the extended form, automatically skipping inactive periods and displaying an explicit closing screen. The extended edition closes after 60 problems. There are no waiting pages and no requirement that the entire class finish before an individual reaches closing. The message confirms recording without offering an ability interpretation.

Basic editions do not return individual profiles. Extended editions allow five private counts only when the person has submitted every active screen and the instructor currently reveals results. Hiding results withdraws that feedback on the next check. A submitted omission counts as a completed screen and contributes zero under the declared rule. An unvisited problem remains pending and prevents the corresponding complete sum.

BEER Lab

Activity complete

Your responses have been recorded. Keep this screen open and wait for the instructor's directions.

If private feedback is enabled, it will appear here after the instructor reveals the results.

Figure 1.7: Closing confirms recording while feedback remains hidden.

1.9 The private report before disclosure

The report shows progress while keeping outcome statistics hidden until the instructor reveals them. First confirm session, language, edition and participant count. Close the access-code panel and consider whether enough observations exist to discuss aggregates without identifying people. In a small class, even an aggregate can disclose individual information when combined with what participants know about one another.

Disclosure should follow the lesson plan and current progress. Avoid treating an early capture as the cohort's final result. Exporting data is distinct from projecting results: private review can be necessary before showing anything to the class. Authentication and dedicated controls separate these operations. Never distribute the report address as a participant entry link.

1.10 Reading the revealed report

The report presents the total and four components, each with its own maximum and completed-observation count. It includes mean, median, sample standard deviation and distribution. Exclusion

BEER Lab

Activity complete

Your responses have been recorded. Keep this screen open and wait for the instructor's directions.

Your descriptive result in this activity

Total: 58 / 60

Letter and number series: 8 / 9

Matrix reasoning: 11 / 11

Verbal reasoning: 16 / 16

Three-dimensional rotation: 23 / 24

Correct counts describe only this form and administration. Explicit omissions and incorrect answers contribute zero; a pending problem is not replaced with an error. Components have different maxima. These are not IQ scores, percentiles or diagnoses.

Figure 1.8: Extended private feedback requires completion and current disclosure.

columns explain which records do not enter a measure. A component can have more observations than the total when some people finish that block but leave other blocks pending. This is expected and should be explained before averages are compared.

Histograms group integer counts into contiguous intervals. Height denotes frequency; intervals do not define psychological categories. A tall bar describes concentration in this session without defining population normality. The item table displays submissions, correct answers, do-not-know choices, omissions and pending records. Its percentage uses submissions as denominator and includes submitted omissions as zeros. Correct-answer keys are absent from the projectable view.

1.11 Downloading and checking data

The custom export includes problem rows for every active period, including pending records, plus five score rows per person. Basic editions therefore produce 21 data rows per participant and extended editions 65, in addition to the header. Three basic positions produce 63 data rows; three extended positions produce 195. These are coverage checks rather than sample sizes for the total score. Measure rows are not additional people.

The private CSV includes prompts and keys together with identifiers, responses, timings, version and hashes. Keep it under restricted access and outside the public repository or public guide folders. Wording and keys support authorized audit but also reconstruct part of the bank. Download only the selected session, check identifiers and store any name correspondence separately. Native wide exports have a different structure; do not join them by row position without verifying keys.

BEER Lab

ICAR instructor report

Form: ICAR-60 · Extended · en

Language version: Original English item files; classroom administration adaptation.

[▶ Class access code](#)

Participants
3

Activities finished
1

Batteries finished
1

Scoreable batteries
1

Reveal aggregates and enable configured feedback

Download this session's responses and scores

Aggregates remain hidden. You can review progress without revealing profiles.

Instrument and provenance

Condon and Revelle (2014), *Intelligence* 43, 52–64, DOI 10.1016/j.intell.2014.01.004. Package provenance: PsychArchives, DOI 10.23668/psycharchives.22167. Materials are separate from public code and subject to their own use terms.

No norms, percentiles, IQ conversions or participant classifications. Recorded time is not a speed score.

Figure 1.9: The instructor chooses when results become visible.

BEER Lab

ICAR instructor report

Form: ICAR-60 · Extended · en

Language version: Original English item files; classroom administration adaptation.

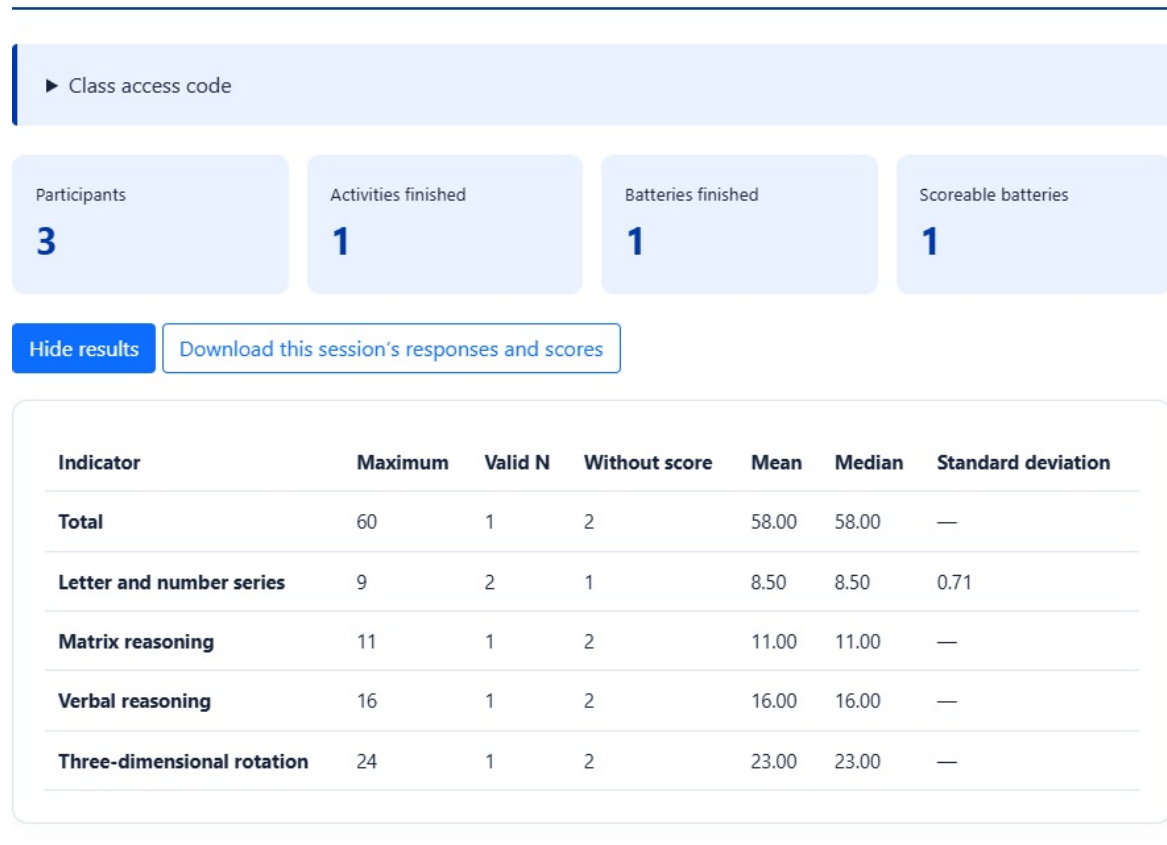


Figure 1.10: Each measure retains its own maximum and observation count.

Total

N = 1 · Frequency



Figure 1.11: Contiguous total-score intervals have frequencies summing to N.

1.12 Lesson planning and pedagogical closing

For a 75-minute class, allow a short procedural introduction, sufficient time for the selected form and at least twenty minutes for interpretation. The basic edition leaves more room for denominator exercises. The extended edition requires local piloting and can be assigned as a supervised preparatory activity, reserving classroom time for discussion. Timing estimates are planning choices rather than normative parameters or automatic program limits.

Close by requesting one descriptive statement supported by the report and one claim the design cannot support. Describing a mean with its N differs from claiming that a language causes lower ability. Assessment of student learning can focus on defending denominators, identifying exclusions and communicating uncertainty. It need not reward test performance. Save the export, manifest and incident notes before ending the administrative session.

Checkpoint

Before starting, check session, form, language and one position per person. Before projection, check that the access-code panel is closed, each measure has its own N, disclosure is intentional and personal identifiers are absent.

1.13 Materials and continuity

The official archive is identified by DOI and hashes. The user accepted its Scientific Use License for their own research and teaching. Prompts, alternatives, keys and figures remain private server inputs. Public software and guides omit the bank. BEER_ICAR_MATERIALS identifies its location; development uses a Git-ignored folder. Without those inputs, session creation fails clearly instead of silently substituting synthetic problems.

Guide figures show operational screens and demonstration reports. Item captures remain private, while worked cases use synthetic counts rather than instrument solutions. Each standalone teaching copy contains PDF, LaTeX and referenced figures. Dropbox copies are checked by SHA-256; compilation auxiliaries and the original archive are excluded. Delivery distinguishes verified local access from public addresses requiring deployment and private provisioning.

Before another administration, check materials, room configuration and closing. Changes to wording, translation or keys require an identified version and the affected checks. Retain provenance and original attribution. When the declared purpose ends, manage material copies under accepted terms and institutional obligations. Participant records and bank materials require separate retention decisions. Keep key-containing exports out of public circulation and document access to private files.

A discussion claim should be reconstructible from its form, cohort and calculation rule. Save an incident note when access, language or interruptions could affect interpretation. The availability of detailed records supports review; it cannot recover information that the administration never collected.

Foundations of measurement and scoring

2.1 From construct to observable record

Cognitive ability is not directly observed in the same way as a marked screen alternative. The available record is a response to a particular problem under specified instructions, language and administration conditions. Interpretation connects that behavior to a broader construct through theory and measurement evidence. That connection does not arise automatically from the existence of a correct key. Begin discussion by separating the administered problem, the observed response and the inference one wants to make.

Performance can depend on prior knowledge, linguistic comprehension, attention, format familiarity, effort and material conditions. Combining tasks broadens content coverage but does not remove every influence. In class, this distinction helps identify the variation of interest and the conditions that should be comparable. A result first describes one administration. Generalization to other tasks, populations or uses requires additional evidence and an explicit account of purpose. Ask students to write that purpose before they inspect results, because an attractive pattern can otherwise lead to a different interpretation after the fact.

2.2 Four problem types

Letter and number series require identifying a regularity and selecting its continuation. A participant represents positions and relationships, maintains a hypothesis and checks it against alternatives. Recognizing an apparent pattern does not establish that it is the only mathematically possible rule. The operational key corresponds to the rule intended by the problem. This distinction separates solving an instrument item from making a universal statement about every conceivable sequence. It also explains why keeping the original alternatives and keys is part of implementation fidelity.

Verbal problems include logic, vocabulary and general knowledge. The label does not mean that every item measures language proficiency alone. Some involve numerical relationships or inferences between statements; others depend on word meanings or learned information. The component should therefore not be treated as a pure skill independent of educational context. When languages are compared, alternatives, meanings and cultural familiarity deserve as much attention as literal prompt translation. A grammatically fluent translation can still change the difficulty of choosing between plausible alternatives.

Matrices display relationships between shapes and ask for a missing part. Solving them requires representing attributes and comparing regularities within an array. Three-dimensional rotation asks

which rendering could represent the same cube after a change of orientation. Relations between faces and symbols matter, together with distinguishing rotation from incompatible alterations. The ICAR bank and a teaching activity inspired by Raven remain different products, although they share some visual reasoning operations. Do not use one instrument's name or validation evidence as a substitute for identifying the other.

2.3 Content coverage and form composition

The 16-item form contains four problems of each type. Its balanced composition supports an initial explanation of components because each block contributes the same possible number of points. The 60-item form is unequal: nine series, sixteen verbal problems, eleven matrices and twenty-four rotations. This matters when interpreting the total. Each correct answer contributes one point, so a component with more items can contribute more points to the combined result.

Giving components equal weight would create a different index from the original total. That transformation can be examined transparently with synthetic data but must not silently replace the defined scoring rule. Keep the original count separate from any index developed for methodological discussion. Likewise, do not compare raw totals with different maxima as though they shared one scale. A percentage describes the correct-answer proportion more conveniently but does not prove equal difficulty or measurement equivalence between forms.

Theoretical foundation

A measurement model connects observed responses to an interpretation of a construct. Correct alternatives support a reproducible scoring rule; they do not by themselves establish that a sum represents general ability, is comparable across languages or supports individual decisions.

2.4 Binary scoring and distinct response states

The basic rule assigns one point to a correct answer and zero to other submitted answers. In this implementation, explicit omission also contributes zero once confirmed. A do-not-know choice is a recorded response with a meaning distinct from selecting an incorrect substantive alternative. Both yield zero under the sum, while remaining distinguishable in the data. None of the alternatives is a keyed option rather than an absence of response and must be evaluated using the item's actual answer mapping.

A pending problem is not an incorrect response. The person's eventual choice is unknown because the screen has not been submitted. The engine preserves that state and withholds a complete sum until every problem belonging to that measure has been submitted. This prevents partial progress from appearing as final performance. State the rule before interpretation because it determines inclusion in each average and helps distinguish interruption, deliberate omission and substantive choice. A common numerical consequence does not make their behavioral interpretations interchangeable.

2.5 The total and components have separate N

The full-form total requires submission of every active problem. A component requires submission of every problem of that type. Someone who completed series but left matrices pending can contribute a series score without contributing a total score. Component N can therefore exceed

total N. This is an intended difference in inclusion sets, rather than inconsistency between report rows. A single participant count at the top of the page cannot replace these measure-specific denominators.

Before comparing rows, identify each maximum, number of complete observations and exclusions. Then ask whether apparent differences reflect performance or the composition of available records. If people experiencing difficulty tend to stop earlier, the final average describes a selected subset. Turning pending responses into zero does not resolve that problem; deleting every partial record also loses information. Document the rule and assess which interpretations remain defensible. The report makes such questions visible rather than supplying an automatic correction for selective completion.

2.6 Center and dispersion

The mean adds included scores and divides by their N. The median orders values and locates their center. They summarize the same distribution in different ways and may coincide without implying symmetry. An isolated mean hides patterns: two classes can have identical averages and markedly different dispersion. Start with observation count and histogram, then use measures of center to summarize the distribution already examined.

The report uses sample standard deviation, with N minus one in the variance denominator. It is not calculated for a single observation. With no observations, neither mean nor median is available. A dash is not zero in these circumstances; it indicates an unavailable or undefined quantity. Explain this using a small example before projecting classroom results. Otherwise a lack of information can be mistaken for uniformity or null performance. Rounding shown in the interface is a display choice; retain underlying counts for reproducible calculation.

2.7 Intervals, percentages and item review

Histogram intervals cover possible integer counts without overlapping. Grouping neighboring values improves readability while sacrificing detail. Their frequencies must add to the N of that measure. If they do not, check coverage, boundaries and filters. Bar color or width supports reading frequency; it is not an ability scale. Do not label intervals low, normal or high without a suitable external reference and validated interpretation.

The item-level correct-answer percentage divides correct responses by submissions. Submitted omissions and do-not-know responses remain zeros, while pending records do not enter that denominator. This answers a specific descriptive question: what fraction of submitted screens contained a correct choice? Dividing by everyone invited answers another question. Dividing only by substantive alternatives changes the target again. Students should write their denominator in words before entering a formula. A correctly executed calculation can still answer the wrong question if its inclusion rule is unexplained.

Percentage is not percentile

A correct-answer percentage compares a numerator with a form's maximum. A percentile places a result within a reference distribution. These are different operations. Seventy-five percent correct does not imply the seventy-fifth percentile or a particular population position. Even a rank within the classroom depends on the people attending and the cohort size. Displaying it graphically would not turn it into a population norm.

The application does not convert scores to IQ, diagnostic categories or personal cutoffs. A published correlation with another instrument is insufficient to build such a conversion. Equating scales requires models, samples and analyses suited to the intended use. Restraint here supports learning rather than preventing it: the class can focus on reproducible operations, clear assumptions and claims its observations actually support. Ask students what reference population and uncertainty would be required before making the stronger claim.

2.8 Timing and administration conditions

The original guidelines describe these forms as power tests: timing is not part of suggested scoring. They mention a possible organizational upper bound for the short form in young adults and recommend piloting other selections. This implementation has no automatic limit. Minutes reserved in the lesson plan organize teaching rather than establishing a normative testing parameter. If an instructor interrupts administration, that decision must be recorded and unsubmitted problems must remain pending.

Recorded screen time describes the interval between presentation and submission. It may include distractions, a background tab or connection interruptions. It is not a pure reasoning-speed measure. Seconds across components are also not directly comparable without considering reading and visual complexity. A study of speed requires an explicit design with timing and exclusion rules set beforehand. Sorting elapsed times in the export does not create that design, and the fastest record should not automatically be interpreted as the most proficient participant.

2.9 What the original sources establish

Condon and Revelle's 2014 article develops and initially validates the resource using four problem types and forms for studying cognitive performance. It supports the identity and composition of the original forms; it does not automatically validate every subsequent translation, interface or use. Materials distributed through PsychArchives allow direct checking of instructions, stimuli and keys. Documentary correspondence is an implementation requirement distinct from local psychometric validation.

Revelle, Dworak and Condon discuss the usefulness of open measures for studying cognitive ability in everyday contexts. Young and Keith examine convergent validity between ICAR-16 and WAIS-IV in a specific sample. These contributions help teach how evidence accumulates across measures. They do not justify assigning each student an individual equivalent from another battery or generalizing without attention to population, language, selection and administration conditions. Describe the question each study addresses rather than using a citation as a blanket guarantee.

Evidence and limits of interpretation

3.1 Reliability, validity and different uses

Reliability concerns aspects of score consistency or precision under a model and specified conditions. Validity concerns evidence and arguments supporting an interpretation for a use. A measure can be consistent without supporting a particular conclusion. A stable sum does not prove that a language comparison is free from cultural differences or that a score is suitable for assigning educational opportunities. These are different claims with different evidentiary needs.

More problems may broaden coverage and reduce some random variation, while longer administration can introduce fatigue, disengagement or differential exposure. The 60-item form is not automatically preferable for every class. Consider time, content and curricular question together. Sixteen problems may be sufficient to teach scoring and denominators. A detailed comparison of short-form components requires caution because each contains only four item responses per person. Distinguish the number of people from the number of problems when discussing precision; increasing one does not reproduce the information supplied by increasing the other.

3.2 Measurement error and uncertainty

An observed score is not a perfectly known personal constant. Attention, motivation and conditions could differ in another administration. A formal measurement model would distinguish systematic and random sources of variation and estimate relevant uncertainty. This teaching interface does not estimate personal intervals or latent parameters. A small difference between two counts therefore does not demonstrate a stable difference between individuals.

Uncertainty about a class average is a separate question. It depends on how the sample was obtained, variation between records and the intended generalization. A convenience cohort does not become representative simply because attendance is large. Ask students to define the population about which they want to speak and explain the inclusion mechanism. They can then decide whether the average is a description of the classroom or whether more data and assumptions are needed. Repeatedly applying a formula cannot compensate for an undefined inferential target.

3.3 Translation, equivalence and declared corrections

The distributed Spanish workbook includes series and verbal wording. Review found that LN.03 repeated another item's sequence; the English source sequence was restored. Two expressions concerning multiplicative age or quantity were also adjusted to preserve the English meaning. Corrections are documented by identifier rather than concealed as though the received file were

flawless. They do not establish that the entire Spanish form has been validated.

Semantic translation is necessary, while measurement equivalence requires more evidence. Familiarity with vocabulary, names, knowledge and alphabets may vary across groups. Comparing Spanish and English averages without randomized assignment and an equivalence study mixes language with participant selection. Identical identifiers and keys do not guarantee identical effective difficulty. Treat Spanish as teaching material with documented provenance and no demonstrated local psychometric validation. The interface's complete translation is a usability requirement, and should not be confused with that separate empirical claim.

3.4 Effort, incentives and prior exposure

Performance requires attempting problems under the stated rules. The absence of payments does not eliminate effort differences, and a reward would not guarantee equivalent motivation. The implementation contains no experimental incentive treatment. When discussing possible motivational mechanisms, distinguish a hypothesis from a result demonstrated by this activity. An omission or a rapid response alone is insufficient evidence of low effort.

Prior bank exposure can create familiarity with problems or keys. Asking about it would be an additional design decision, not information already collected here. For an ordinary class, avoid distributing solutions before administration and do not reuse the bank as though every exposure measured performance without learning. Worked examples in this guide use synthetic numerical records to teach operations without disclosing instrument solutions. This separation protects both the teaching opportunity and the interpretability of later administrations.

3.5 Connections to behavioral economics

Information-processing capacity may help explain decisions, but it is not a preference. A count of correct answers does not directly measure patience, risk aversion, cooperation, self-control or motivation. People with the same score may make different economic choices for legitimate reasons involving preferences, information and context. An error in an economic game also does not, by itself, establish lower cognitive ability.

A useful curricular extension discusses variables that could confound an association between performance and decisions. Prior education, language, technological familiarity and sample selection are conceptual candidates, not automatically observed variables in this application. Studying a causal relationship requires a design addressing that question. Combining classroom activities is valuable when it makes hypotheses and measurement explicit, rather than labeling participants or building retrospective explanations from a small set of observations.

Feedback script

"This report describes responses in one session under an identified form and set of conditions. Before explaining a difference, check the maximum, N, pending records, language and prior exposure. A count informs us about a task; a broader interpretation requires additional arguments."

3.6 Comparing forms and conditions

The basic edition fits a short experience followed by substantial discussion of scoring. The extended edition adds content and allows examination of unequal component contributions. It is not a second level unlocked after completing basic. They are alternative forms for a single administration. When groups use different forms, retain separate outcomes, maxima and denominators.

A didactic percentage comparison can show why an arithmetic transformation does not establish equivalence. If the longer set adds different difficulty and composition, the proportion can change without a change in underlying ability. Before deciding which group performed better, require an operational definition and examine whether the design supports it. The most useful educational conclusion may be that additional information is needed. Students should be able to explain exactly which missing comparison or assumption prevents the inference.

Order, fatigue and assignment

Fixed order makes the current block predictable and helps explain component completion. It also links problem type with position because rotations come last. A position-related decline may reflect difficulty, composition, fatigue or selection of those continuing. The activity does not identify those mechanisms separately. Acknowledge that limit instead of presenting a temporal curve as direct evidence of exhaustion.

An order study would require comparable conditions, participant assignment and preservation of the sequence actually received. That study is outside the current four configurations. It can be proposed as a design exercise, clearly distinguished from implemented behavior. The same applies to incentives, permitted tools and time limits: they are substantive administration changes rather than hidden options already available in the panel. A proposed improvement should name its question, manipulation, outcome and comparison before adding interface complexity.

3.7 A learning rubric

Assess the ability to reconstruct a measure: identify included observations, defend numerator and denominator, calculate accurately and recognize an unsupported inference. Request a short note containing a table and an interpretive paragraph. An excellent response may explain that a requested comparison is not identified by available data and propose what would improve it. That conclusion must be grounded in instrument and record facts rather than a general statement that all measurement is imperfect.

A rubric can award points for traceability, calculation, interpretation and communication without grading a student's ICAR performance. Distinguish arithmetic from conceptual mistakes: dividing a four-item component by sixty differs from adding incorrectly. Use the following synthetic cases so everyone first works with the same records. Discussion of the class distribution can then apply an already understood method rather than turning classmates into examples of personal success or failure.

Attention

Language, order, presentation and omission adaptations are declared. Implementation preserves original form inventories and keys, while establishing neither local psychometric equivalence nor diagnostic use or individual IQ conversion.

Worked cases for interpretation practice

What participants will see

First solve cases with synthetic data. Identify form, problem type, key, response status, inclusion rule, numerator and denominator before calculating.

Scoring and missing data

Case 1. A complete short form

A person submits all sixteen problems and answers ten correctly. Six substantive choices are incorrect and none is pending. The score is ten and the correct-answer proportion is ten divided by sixteen, or 62.5 percent. The denominator is the form maximum rather than the number of alternatives in each problem. Averaging the numeric button positions would not score performance.

This record enters the total-score average because every problem was submitted. It does not establish a population percentile. Communicate form, language, count and maximum, then identify what stronger interpretation would require a reference distribution and measurement evidence.

Case 2. Reconstructing an extended total

A complete record contains six correct series answers out of nine, eleven verbal answers out of sixteen, seven matrices out of eleven and fifteen rotations out of twenty-four. Adding components gives thirty-nine correct answers out of sixty, or 65 percent. Component maxima must also sum to sixty.

This check can detect transcription and denominator errors. The unweighted mean of component percentages would be a different index because components contain unequal numbers of problems. Preserve the four fractions when discussing composition. Do not silently substitute an equal-component index for the original item-weighted total. A traceable report identifies both the arithmetic operation and its intended interpretation.

Case 3. A submitted omission

A person answers fifteen short-form problems correctly and explicitly omits the final problem. Every screen has been submitted. The sum is fifteen out of sixteen because confirmed omission

contributes zero under the stated rule. Completion and absence of omissions are different properties. The CSV should contain one omitted row, no pending rows and fifteen correct answers. Omission must not be described as the bank's substantive alternative zero: zero is an interface control. The instructor can explain its arithmetic effect without inferring why the participant used it. The data record the action but do not identify an underlying motive.

Case 4. The last problem remains pending

Change the previous case: the person has not yet visited the last screen. Fifteen observed answers are correct and one problem is pending. A complete total is unavailable. Reporting fifteen out of sixteen as final would treat the pending problem as incorrect, which this implementation does not do.

The difference depends on response state rather than the fifteen correct answers. Completely submitted components can remain scoreable. When reviewing an early report, first read progress and pending counts. Wait for completion or declare the selection rule before describing a final average. Do not interpret the dash as evidence of zero performance.

Case 5. Do-not-know and substantive errors

A complete short form contains eight correct answers, four incorrect substantive alternatives and four do-not-know choices. The sum is eight and the percentage fifty. There are sixteen submissions and no explicit omissions. Selecting twelve substantive alternatives does not change the form maximum.

Eight divided by twelve describes a different quantity that excludes do-not-know responses; it is not the implemented score. Preserving response kinds allows transparent examination of the difference. Do not automatically interpret do-not-know as low effort or an incorrect alternative as confident error, because confidence is not elicited. Shared scoring consequences do not make response meanings identical.

Case 6. One completed component

An extended-form participant submits all nine series problems and answers seven correctly, leaving the other fifty-one problems pending. The series score is seven out of nine; the sixty-item total is unavailable. This person contributes one series observation and no total-score observation.

Series N of one and total N of zero are therefore correct. Filling remaining components with zeros to align rows would alter the rule. Discarding the observed series would also lose information. What can be compared depends on the question and the process producing partial progress. State that process as far as it is known rather than inventing a reason for stopping.

Averages, dispersion and charts

Case 7. Different denominators across rows

Three extended positions have total scores forty, twenty and pending. The total average is thirty, using two observations. Their completed series scores are six, three and eight, so the series average is seventeen divided by three, approximately 5.67, using three observations. Respective maxima are sixty and nine.

Comparing these averages without maxima and N is meaningless. Even percentage comparisons must acknowledge different people in the inclusion sets. A question requiring a common sample needs an additional, explicitly defined filter. The standard report preserves each measure's information under its own completion rule instead of concealing partial records to force equal N.

Case 8. Mean, median and dispersion

Consider complete short-form scores two, three, three and twelve. Their mean is five and median three. Squared deviations from the mean sum to sixty-six. Dividing by three gives sample variance twenty-two; its square root is approximately 4.69.

The value twelve shifts the mean relative to the median. It need not be removed merely because it differs; it is within the admissible range. Exclusion requires a documented reason and rule. A histogram helps explain why one measure of center cannot characterize the distribution. Keep the four observations visible while students reconstruct the calculation, then ask which summary responds to their descriptive question.

Case 9. One observation only

Only one person finishes the short form, scoring twelve. Mean and median are twelve, but sample standard deviation is undefined. The report should display a dash for dispersion. Showing zero would suggest an estimate unavailable from the sample formula when N equals one.

There is also a communication issue: an aggregate of one can disclose that person's result if classmates know who finished. The instructor may decide against projection within the teaching procedure. That visibility decision does not alter the data or make the record missing for authorized private review. Distinguish availability, calculation and disclosure when explaining the report.

Case 10. Checking a histogram

For the short form, use integer intervals zero through two, three through five, six through eight, nine through eleven, twelve through fourteen and fifteen through sixteen. Suppose their frequencies are one, zero, two, one, two and zero. They sum to six observations, and each admissible integer belongs to exactly one interval.

This check does not require identifying people. If the table reports N of seven, investigate the discrepancy rather than changing a bar manually. Review filters, records and boundaries. The intervals describe frequencies, not personal ability categories. A complete audit verifies both that counts add correctly and that every possible value has one destination.

Case 11. The item-level correct percentage

Among ten positions, one problem has three correct answers, two substantive errors, one do-not-know response, one omission and three pending records. Seven screens were submitted. The displayed percentage is three divided by seven, approximately 42.86. Changing the denominator does not change the correct count, but changes the question.

Dividing by ten gives thirty percent of invited positions with a correct answer. Dividing by five gives sixty percent among substantive choices. All can be described, but only the first calculation matches the report's item indicator. An adequate explanation names included and excluded states for every denominator and does not present the three percentages interchangeably.

Case 12. Component weighting

A complete extended form has nine correct series responses, sixteen verbal responses and zero correct matrices or rotations. Its original total is twenty-five out of sixty, approximately 41.67 percent. Averaging the four component percentages equally would give fifty percent because two components are perfect and two zero.

This discrepancy illustrates weighting rather than an arithmetic mistake. The original total gives each problem one point; the alternative index gives each component equal weight. A methodological assignment can calculate both, provided it identifies the transformation and justifies its weights. It must not relabel the second index as the original ICAR total or compare them without explaining their different construction.

Case 13. Equal percentages in different forms

One person scores twelve out of sixteen and another forty-five out of sixty. Both proportions are seventy-five percent. This does not prove equal ability or equal administration difficulty. The forms differ in content composition and number of problems, and may have been administered under different conditions.

The defensible descriptive conclusion is that observed proportions coincide within their respective forms. A broader comparison requires justified equivalence and attention to uncertainty. Do not use coinciding percentages as an improvised conversion table. This exercise separates arithmetic normalization from measurement equating, and encourages students to specify what would be needed before claiming interchangeability.

Case 14. Percentage does not determine rank

Everyone in a hypothetical class scores twelve out of sixteen. Each has seventy-five percent correct, but that does not establish the seventy-fifth percentile. Classroom records are tied and no external population or reference distribution has been specified.

First explain the intended comparison and how ties would be handled. Then recognize that an internal classification would still not establish a population norm. The example separates a correct-answer proportion, ordering observations and normative reference. The application preserves counts and descriptive percentages without generating personal percentile ranks. Asking students to identify the missing reference is more informative than inventing a rank simply because a numerical percentage exists.

Design, selection and traceability

Case 15. Selection through completion

Of ten positions, three finish first with short-form scores fourteen, thirteen and fifteen. The available average is fourteen and N three. Seven positions remain pending, so the final average of all ten is unknown. If those seven hypothetically finished with three correct answers each, the average would become 6.3. This is an illustration, not an imputation.

The difference shows why an early average must not be presented as final classroom performance. It also does not establish that unfinished participants will score less. The conclusion concerns missing information and selection. Preserve observation time and N with any capture used in discussion, and distinguish a hypothetical sensitivity example from observed data.

Case 16. Interrupting the long form

A participant submits forty problems, answering twenty-five correctly, and leaves twenty pending after an interruption. No complete total is available. Twenty-five divided by forty describes observed submissions, while dividing by sixty implicitly treats pending responses as zeros. Neither should be labeled the implemented final score.

Entirely submitted components remain usable. Record the interruption and administrative stopping criterion without automatically attributing missing answers to difficulty. Comparing interrupted and complete participants requires explaining selection. Screen timings alone cannot determine why the interruption occurred. This case demonstrates that a detailed export does not eliminate the need for substantive knowledge of how the records were produced.

Case 17. Administering the short subset first

An instructor proposes ICAR-16 followed by ICAR-60 for the same people to measure learning. The second administration contains previously seen problems. Any difference mixes exposure, changed content, length and administration conditions. It is not, by itself, an estimate of learning or an independent form comparison.

One curricular alternative assigns a single form per person and discusses comparison limits. Another designs a study specifically about exposure with appropriate conditions. That study is not implemented by the current four rooms. This operational recommendation protects interpretation in an ordinary class while leaving room for students to propose better research designs with explicit counterfactual comparisons.

Case 18. Comparing languages without assignment

Two groups have extended means forty-eight and forty-two, one answering Spanish and the other English. The raw difference is six correct answers. It cannot be attributed to language without considering who selected each version, prior education, familiarity, administration conditions and translation equivalence.

An adequate report presents means, maxima, N, form and inclusion rules before advancing a hypothesis. For a causal question, propose language assignment and specify a population able to answer in both languages. A descriptive contrast may motivate that study but cannot replace its design. Identical source identifiers and keys support traceability without proving that both groups face equivalent effective difficulty.

Case 19. Counting CSV rows

Three basic positions produce forty-eight item rows and fifteen measure rows, totaling sixty-three data rows. Three extended positions produce one hundred eighty item rows and fifteen measure rows, totaling one hundred ninety-five. The header is not an observation. Five measure rows per person represent the total and four components.

An untouched position must retain its active item rows as pending. If it disappears, the instructor cannot audit the session universe. Count unique participant identifiers and apply the relevant completion rule when computing person N; never treat all export rows as independent participants. Row counts test coverage rather than determine the sample size for every analysis.

Case 20. Changing materials after session creation

A session is created with an identified package version. The JSON source is later corrected for future administrations. The existing session retains the wording, keys, order and hashes frozen at creation. It must not be silently reinterpreted as though it used the revised version.

Create a new session for revised material and document the change. Preserve graphical files required by existing sessions; replacing them under the same path creates an integrity discrepancy. Reproducibility requires records together with the material and rules that generated them. A current source file alone cannot reconstruct every historical administration if earlier versions were not retained and identified.

Case 21. Downloading a specified cohort

An administrator has two sessions with the same language and edition, each containing three positions. A download requested from one report must contain only that session identifier, rather than mixing all six people. Matching configurations do not make the records a single cohort.

Check the session identifier in every row and compare the expected position count. Combining cohorts later is an explicit analytical operation that must preserve provenance and examine administration differences. Do not remove the session column merely to simplify a table before verifying its contents. Isolation protects the meaning of the download even when an administrator has access to multiple sessions.

Case 22. Revealing and hiding again

An extended participant finishes while results remain hidden. Closing confirms recording but does not display the five counts. Disclosure enables their private display; hiding withdraws them on the next check. Basic editions do not provide that individual feedback.

Visibility does not change responses or scores and is different from deleting records. Separate recording, calculation and communication when explaining the mechanism. A class discussion may require a private export while projected results remain hidden. Those choices are compatible when made from the authorized panel. Check current disclosure rather than assuming that a completed form permanently entitles a browser to keep requesting visible feedback.

Checkpoint

Check session, language, edition and access. Review N per measure, choose disclosure and download only the intended cohort. Do not convert correct-answer counts to IQ or classify people.

References and source materials

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52-64. <https://doi.org/10.1016/j.intell.2014.01.004>

International Cognitive Ability Resource Team. (2026). ICAR material archive; Guidelines, original item files, scoring keys and Spanish workbook. PsychArchives. <https://doi.org/10.23668/psycharchives.22167>

Revelle, W., Dworak, E. M., & Condon, D. M. (2020). Cognitive Ability in Everyday Life: The Utility of Open-Source Measures. *Current Directions in Psychological Science*, 29(4), 358-363. <https://doi.org/10.1177/0963721420922178>

Young, S. R., & Keith, T. Z. (2020). An Examination of the Convergent Validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, 38(8), 1052-1059. <https://doi.org/10.1177/0734282920943455>

PsychArchives. Scientific Use License v1. Terms accepted for the declared teaching purpose. <https://doi.org/10.23668/psycharchives.4988>

American Economic Association. JEL Classification System. A22, C83, C91, D91 and I21 identify teaching, survey methods, individual experiments, cognitive factors and education. <https://www.aeaweb.org/econlit/jelCodes.php?view=jel>