



The application of validity theory in vocational awarding organisations

Paul E. Newton

Federation of Awarding Bodies Annual Conference
20-21 October 2016

Validity is...

“Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests” (AERA, APA and NCME, 2014, p.11)

“Validity is a conceptual animal at war with itself.” (Baker, 2013)

Validity is...

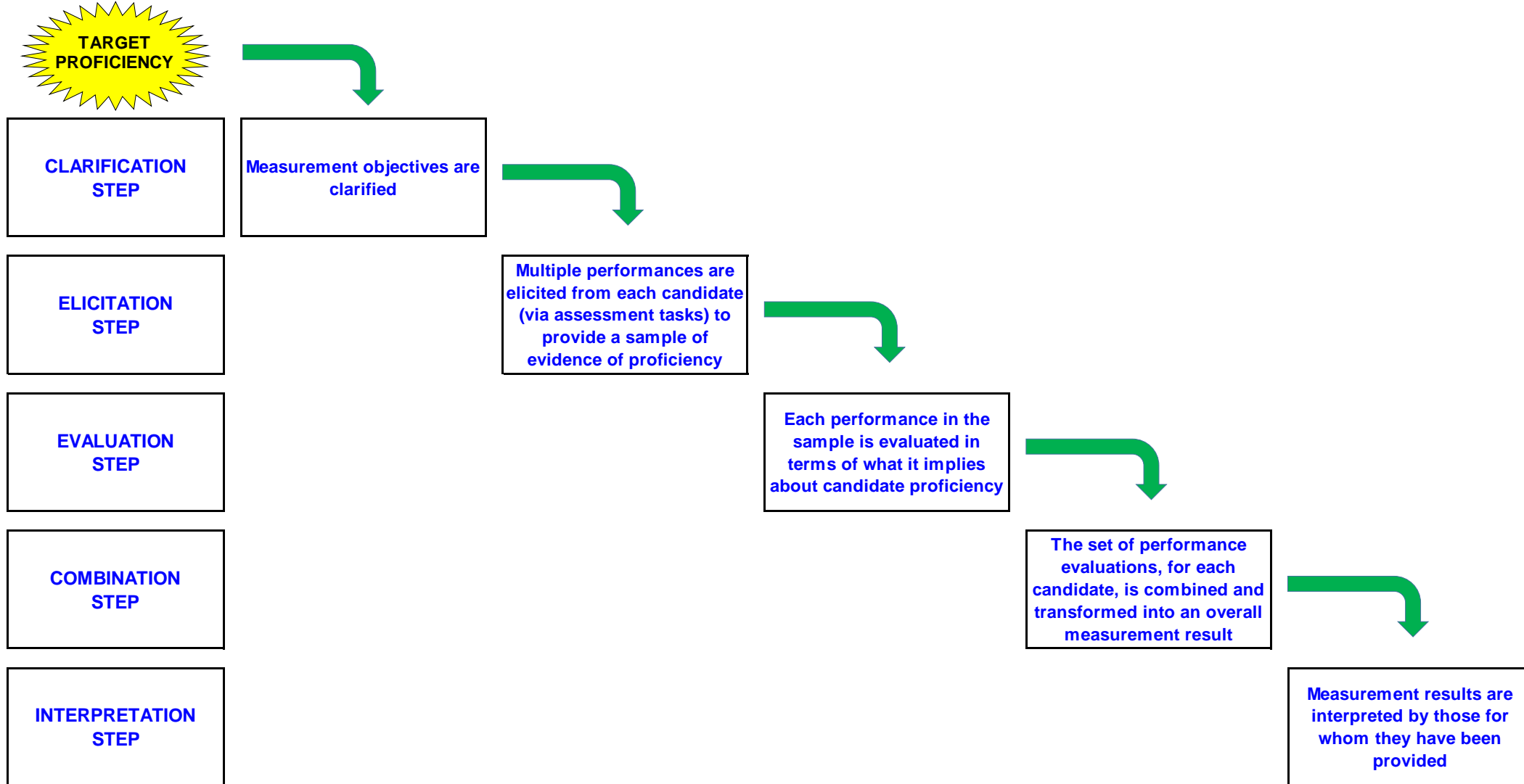
- The validity of a **particular qualification** is
 - the degree to which it is possible to **measure**
 - whatever that qualification **needs** to measure
 - by implementing its assessment **procedure**

What a qualification 'needs' to measure

- The learning outcomes that need to be measured are determined by the **intended use** of certificated results.
 - e.g. if the primary use of outcomes from a plumbing qualification is to allocate (or withhold) a licence-to-practise, then we need to measure the learning outcomes that are fundamental to safe and competent plumbing.
- The intended use of certificated results therefore determines the **target proficiency** – the overarching 'thing' that the qualification needs to measure.
 - e.g. 'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'
- The target proficiency then becomes the reference point for all future **validation research**
 - which investigates the degree to which it's possible to measure the target proficiency, by implementing the qualification's assessment procedure

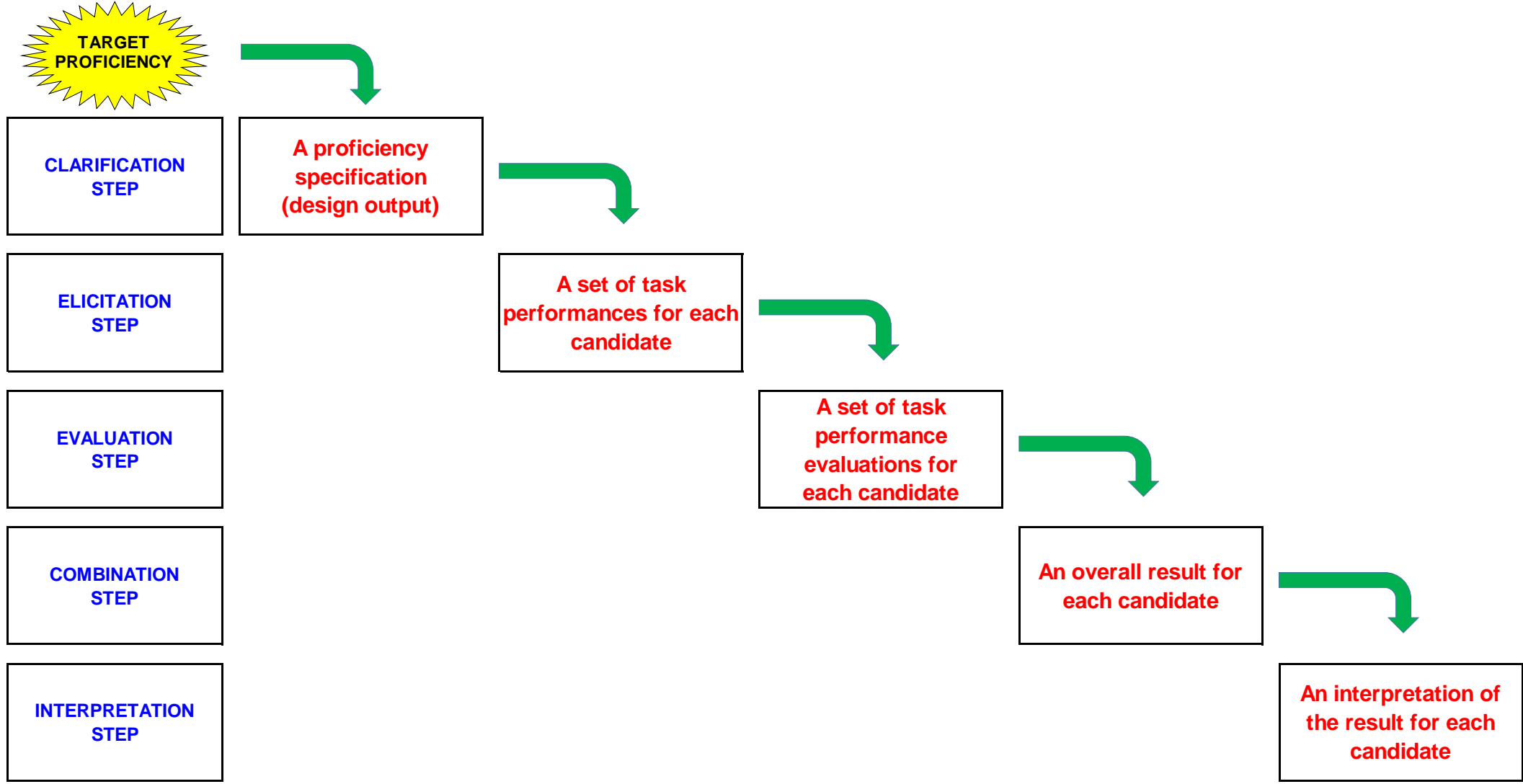
'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'

Logical steps in the assessment process



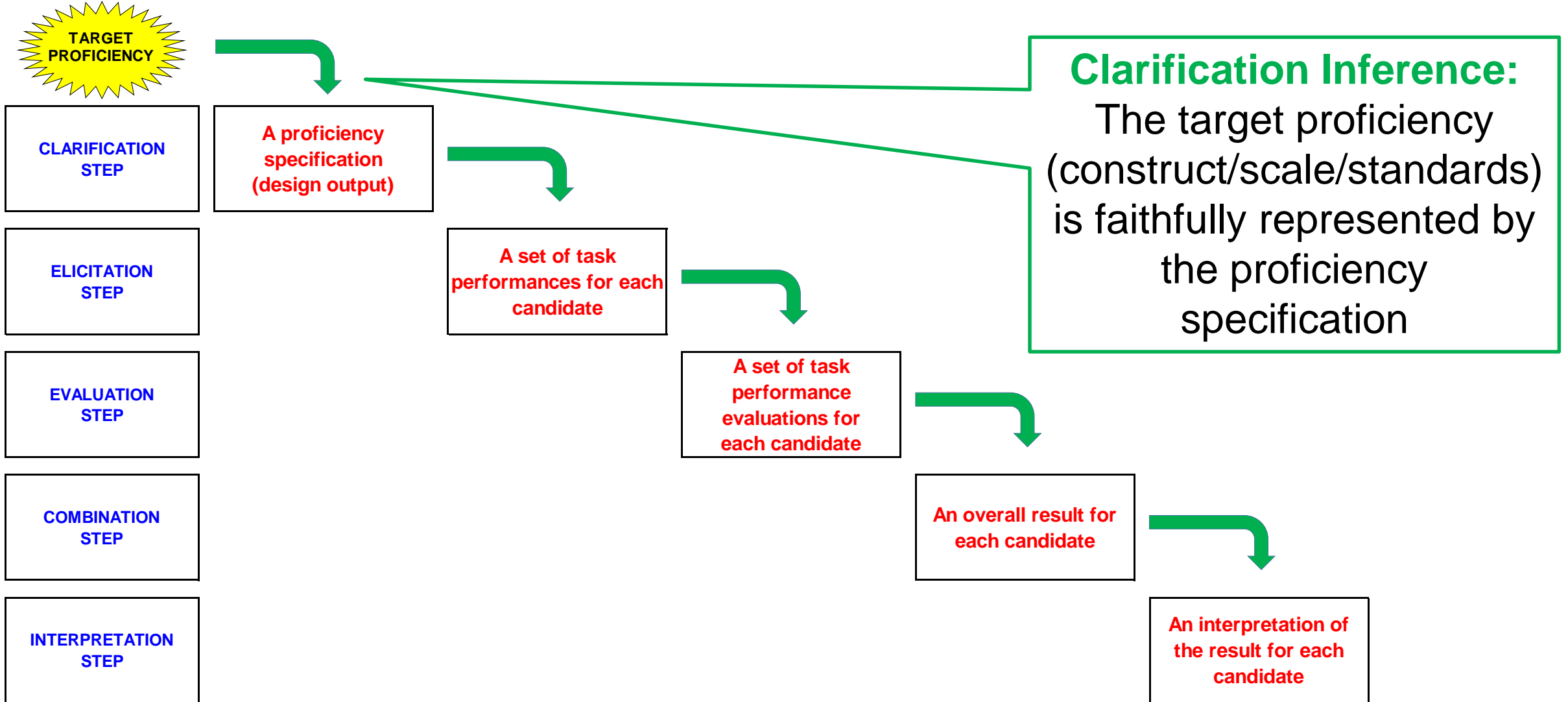
'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'

Final output from each step

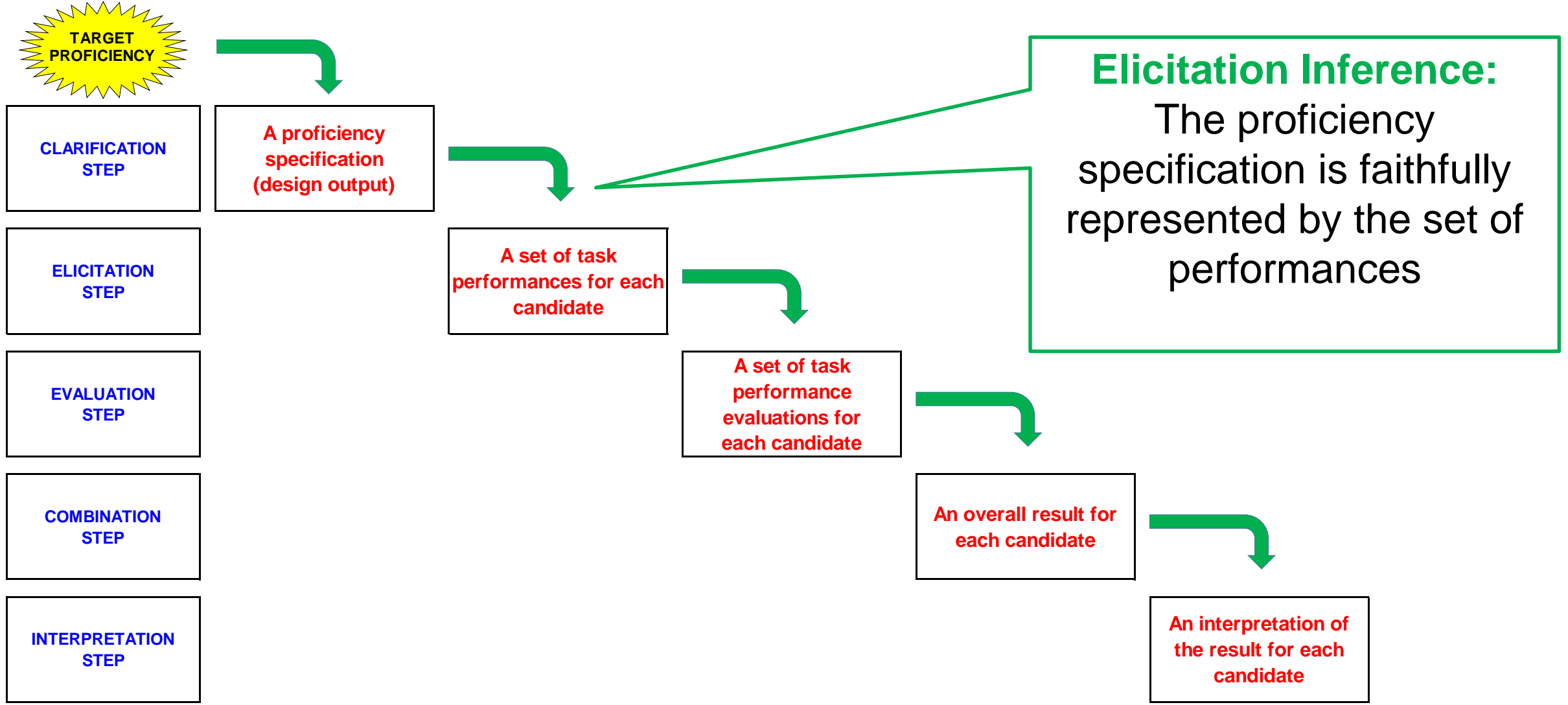


'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'

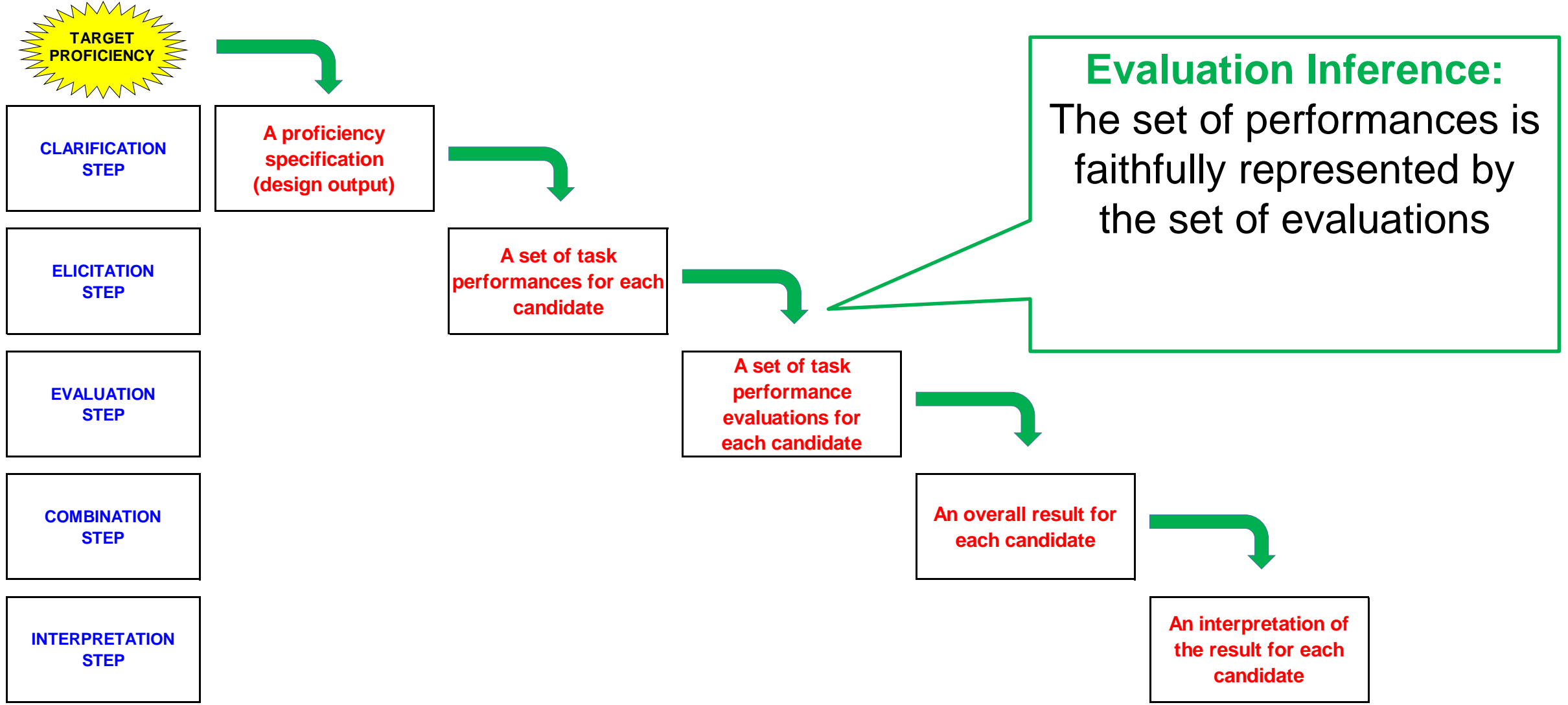
Critical inferences for each step



'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



CLARIFICATION
STEP

ELICITATION
STEP

EVALUATION
STEP

COMBINATION
STEP

INTERPRETATION
STEP



A proficiency
specification
(design output)



A set of task
performances for each
candidate



A set of task
performance
evaluations for
each candidate



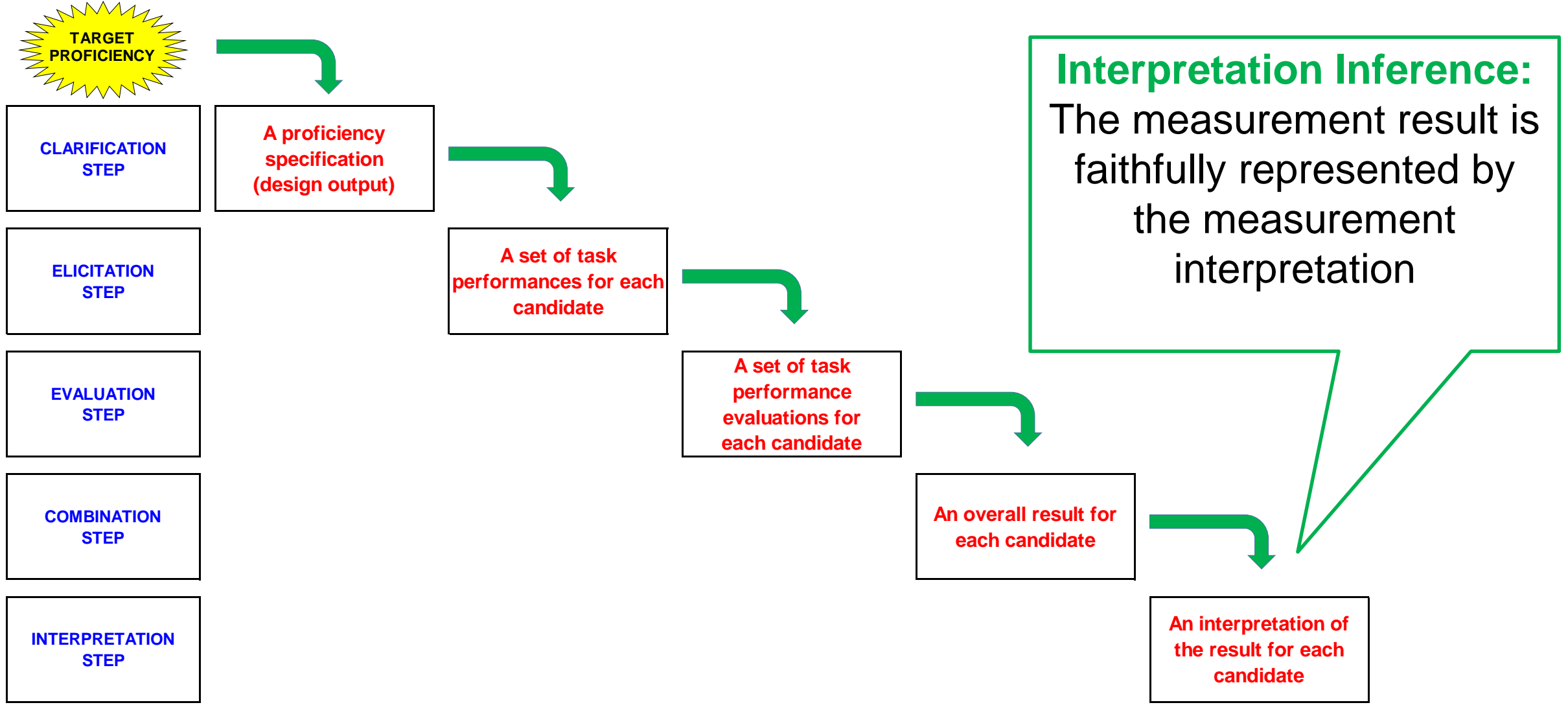
An overall result for
each candidate



An interpretation of
the result for each
candidate

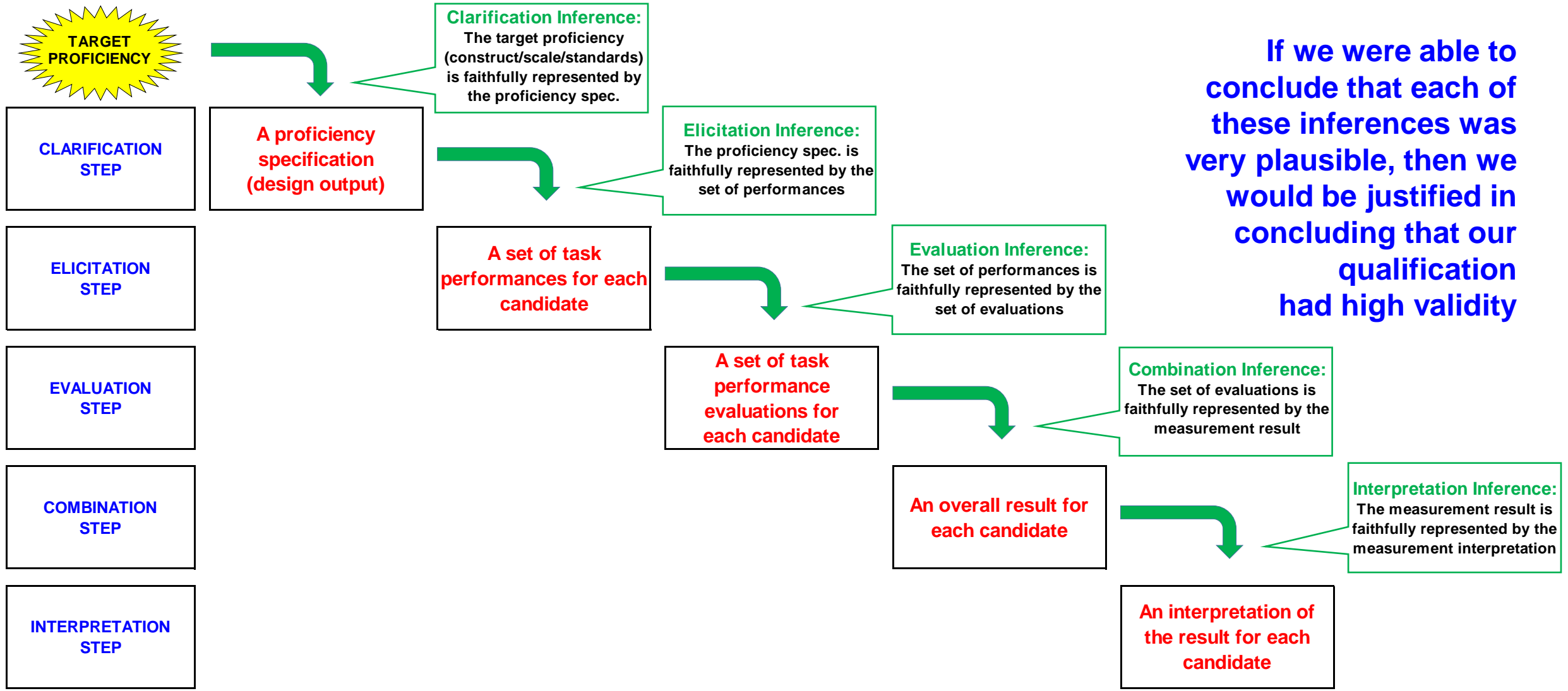
Combination Inference:
The set of evaluations is
faithfully represented by
the measurement result

'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



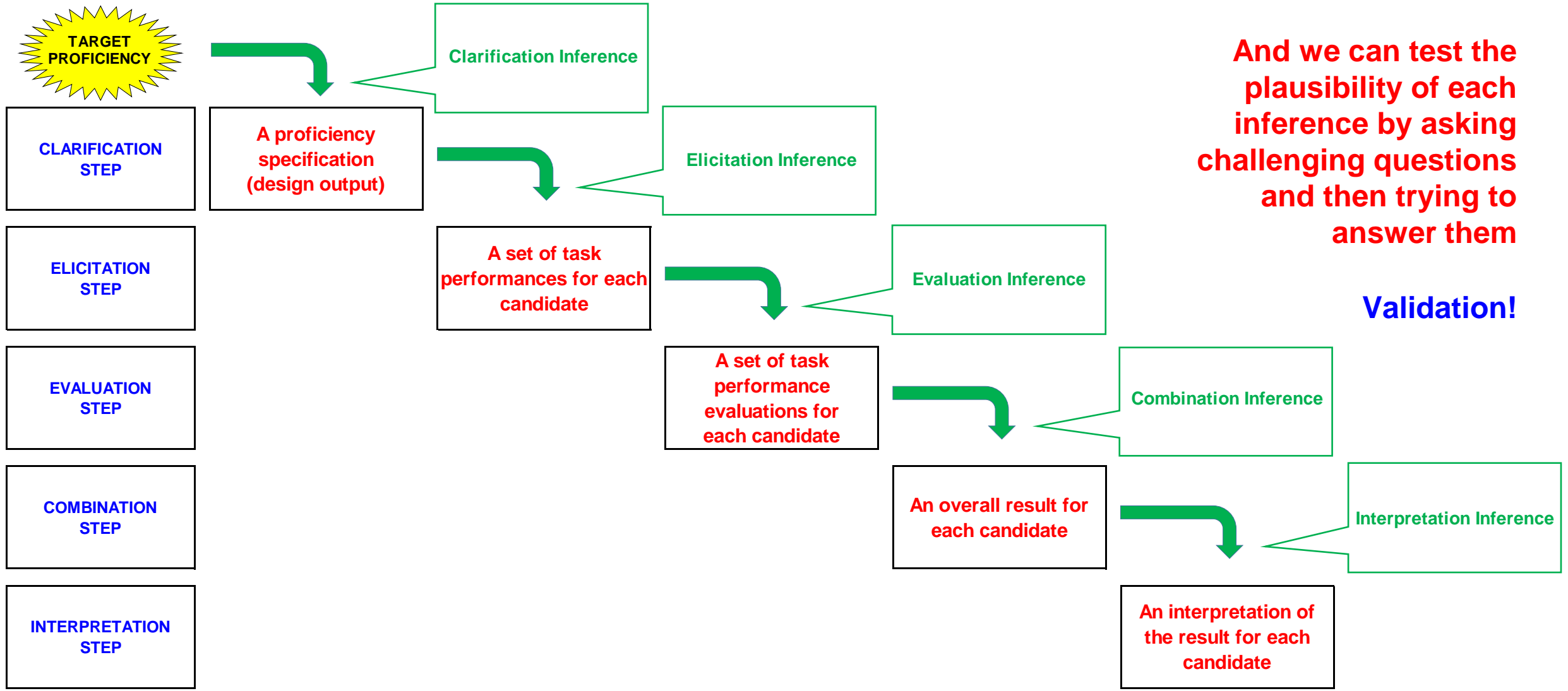
'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'

Critical inferences for each step

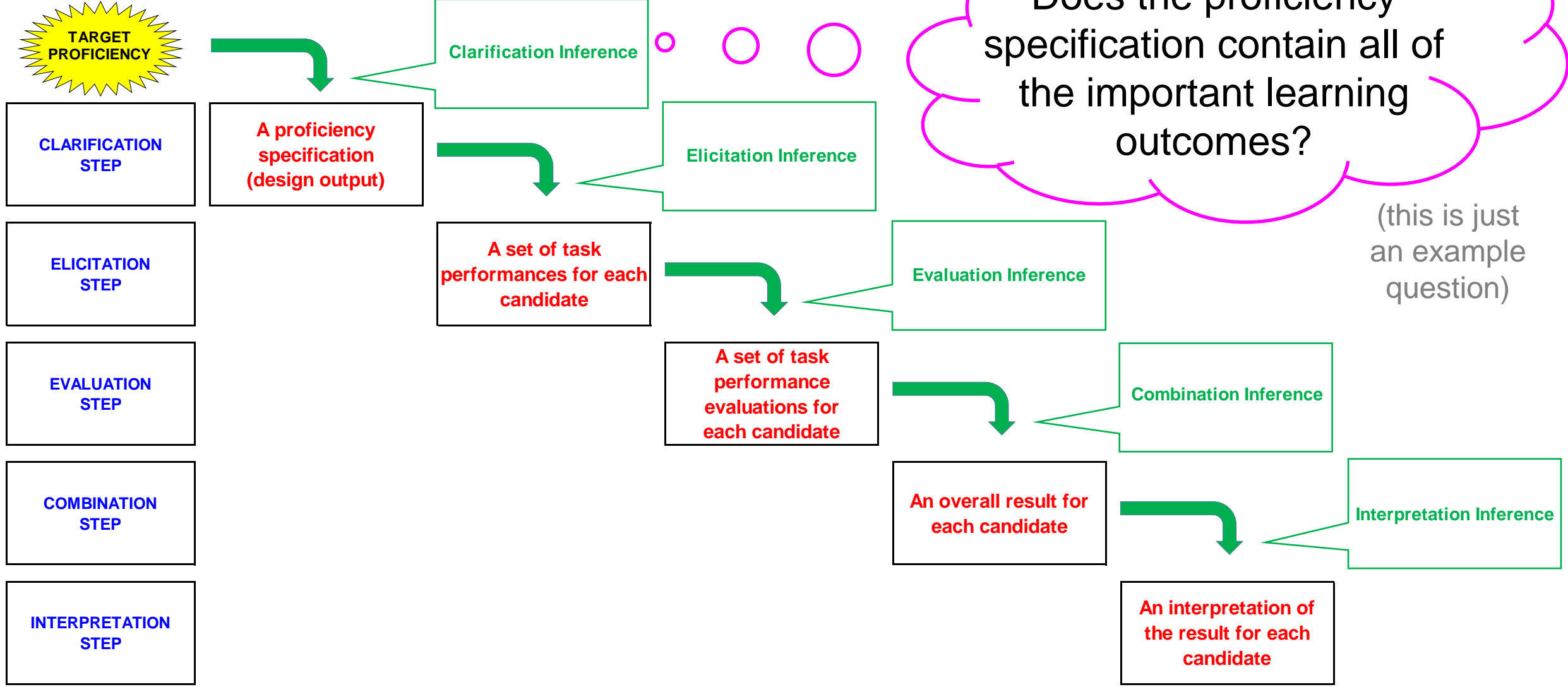


If we were able to conclude that each of these inferences was very plausible, then we would be justified in concluding that our qualification had high validity

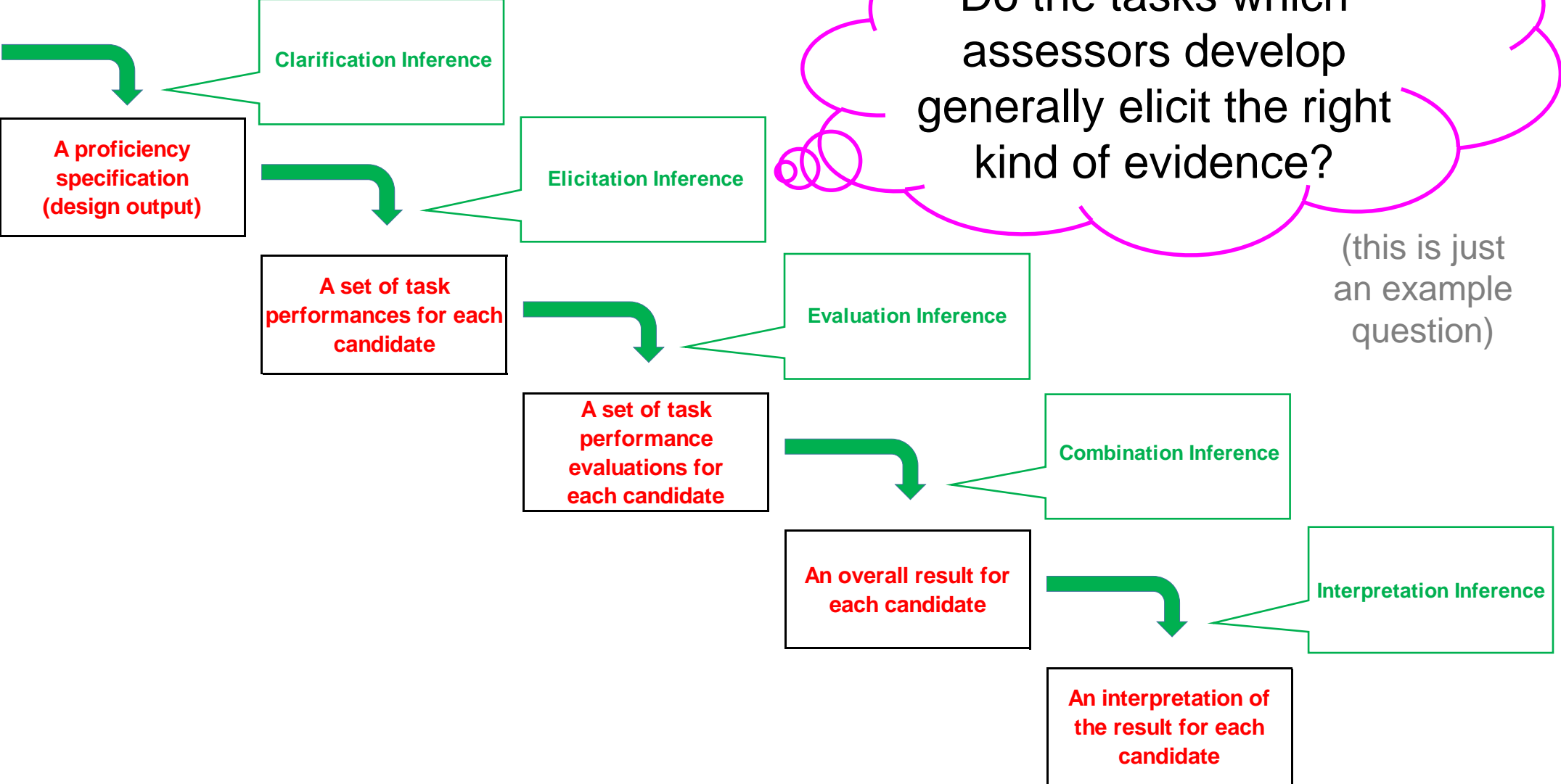
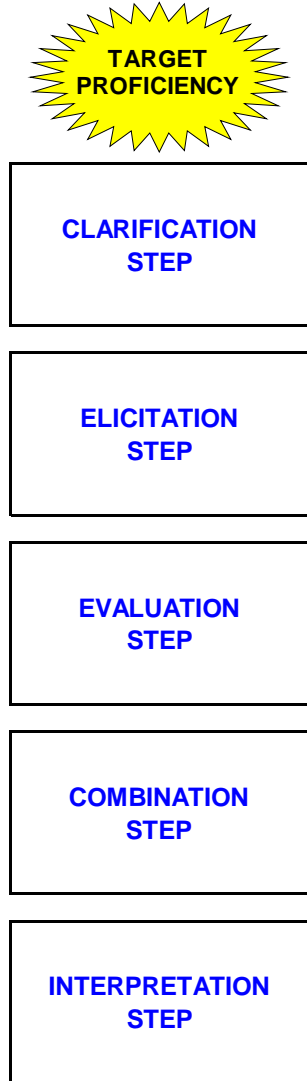
'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'

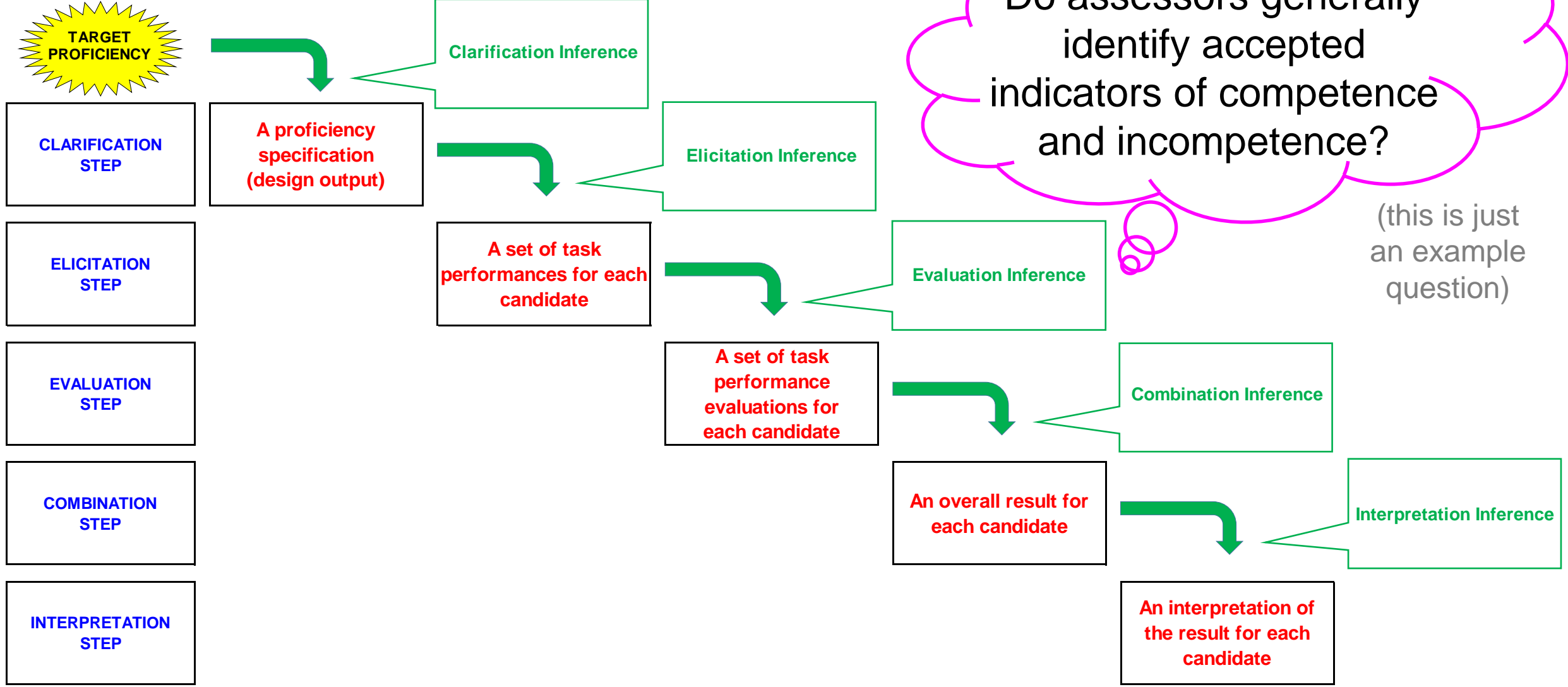


'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'

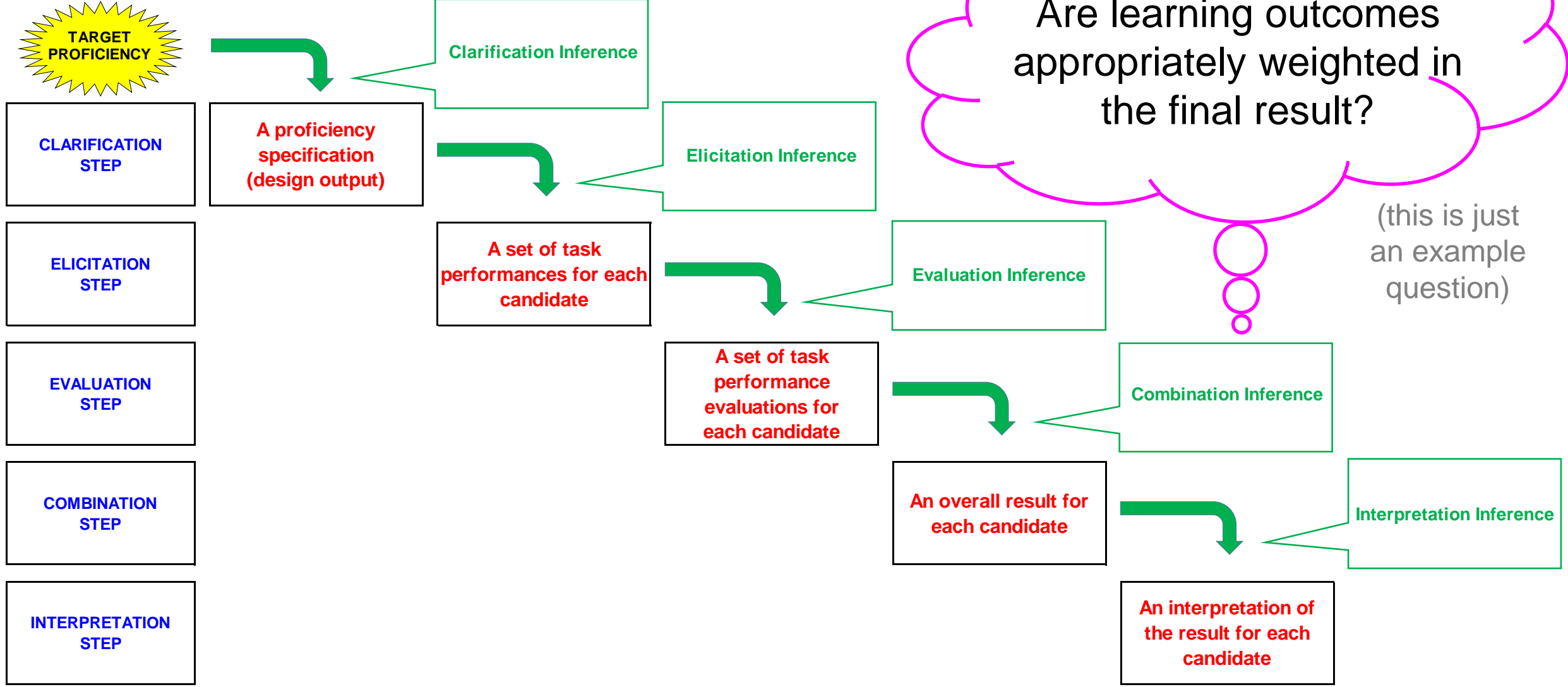


Do the tasks which assessors develop generally elicit the right kind of evidence?
(this is just an example question)

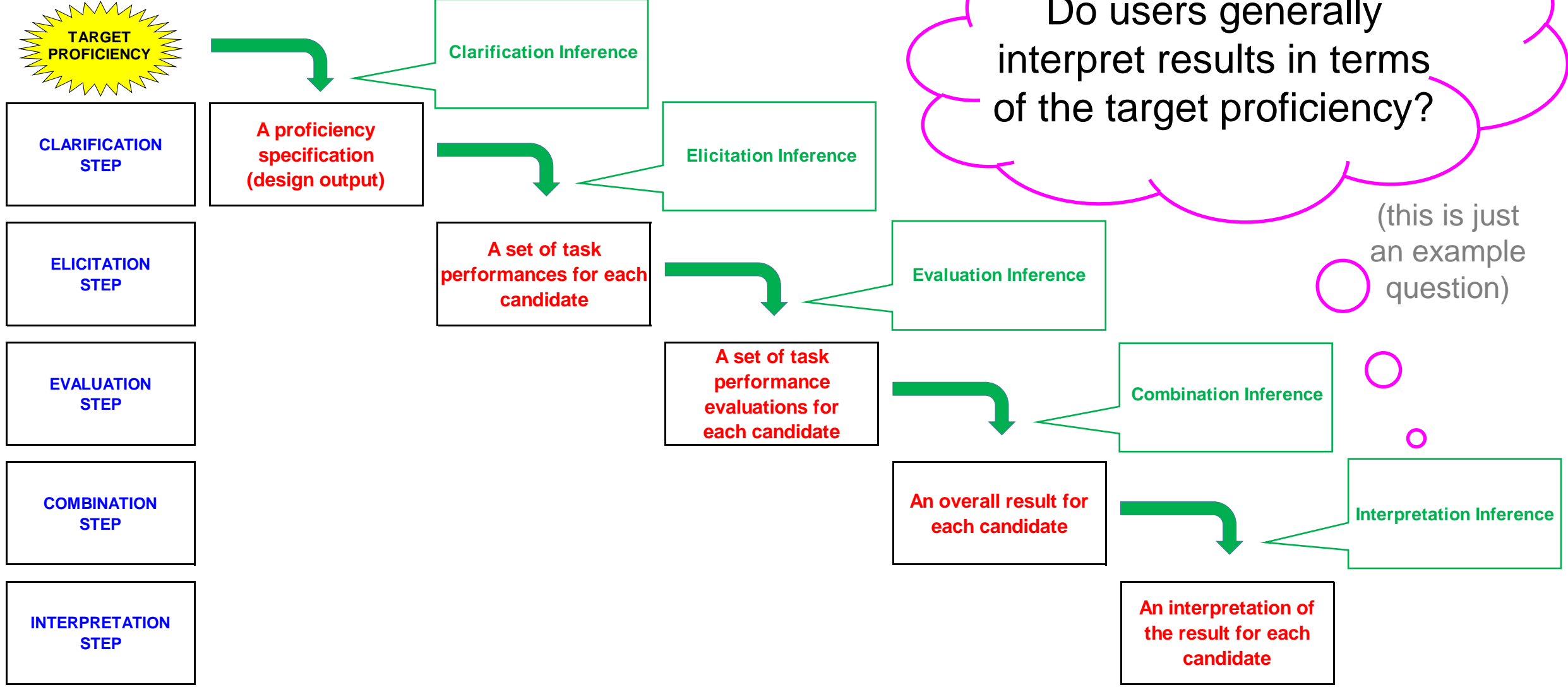
'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



'proficiency in domestic plumbing and heating (gas fired water and central heating appliances)'



Do users generally interpret results in terms of the target proficiency?

(this is just an example question)

TARGET PROFICIENCY

CLARIFICATION STEP

ELICITATION STEP

EVALUATION STEP

COMBINATION STEP

INTERPRETATION STEP

Clarification Inference

Elicitation Inference

Evaluation Inference

Combination Inference

Interpretation Inference

A proficiency specification (design output)

A set of task performances for each candidate

A set of task performance evaluations for each candidate

An overall result for each candidate

An interpretation of the result for each candidate

Example validation question for Inference 1: Does the proficiency specification contain all of the important learning outcomes?

■ Sub-question(s)

- is the 'constellation' of learning outcomes which has been specified for the qualification as relevant to, and as representative of, the target proficiency as possible?

(this is just an example sub-question)

■ Potential validation evidence/analysis

- consultation exercise with users and stakeholders which led to the original identification of the target proficiency
- follow-up research with different users and stakeholders

(this are just example sources)

Example validation question for Inference 2: Do the tasks which assessors develop generally elicit the right kind of evidence?

■ Sub-question(s)

- are formal credentials required of the assessors who develop the tasks?
- are assessors required to undertake qualification-specific training in task development?
- are there parameters for developing assessment tasks and guidance on developing them?
- how is task development quality assured?

■ Potential validation evidence/analysis

- declaration that all current assessors possess (assessor) qualifications
- records of attendance at, and success during, training events
- document files, ideally annotated to explain rationale for decisions and suitability of guidance
- task development procedure files and monitoring reports
- assessment 'task content' research
- candidate 'response process' research

Example validation question for Inference 3: Do assessors generally identify accepted indicators of competence and incompetence?

■ Sub-question(s)

- are formal credentials required of the assessors who evaluate the task performances?
- are assessors required to undertake qualification-specific training in performance evaluation?
- are there parameters for developing assessment criteria and guidance on developing them?
- how is performance evaluation quality assured?

■ Potential validation evidence/analysis

- declaration that all current assessors possess (assessor) qualifications
- records of attendance at, and success during, training events
- document files, ideally annotated to explain rationale for decisions and suitability of guidance
- performance evaluation procedure files and monitoring reports
- assessor 'evaluation process' research
- assessor consistency research

Example validation question for Inference 4: Are learning outcomes appropriately weighted in the final result?

■ Sub-question(s)

- are certain learning outcomes more fundamental to the target proficiency than others, and therefore deserve a higher weighting in the overall judgement of proficiency?
- which, if any, learning outcomes are necessary requirements for passing, or for achieving certain grades?

■ Potential validation evidence/analysis

- detailed scrutiny of specified learning outcomes in parallel with detailed scrutiny of target proficiency
- aggregation model, ideally annotated to explain rationale for decisions
- research into achieved vs. intended weighting of components
- assessor 'aggregation process' research

Example validation question for Inference 5: Do users generally interpret results in terms of the target proficiency?

■ Sub-question(s)

- what information is provided to users (e.g. employers) on how to interpret qualification results?

■ Potential validation evidence/analysis

- document files, ideally annotated to explain suitability of guidance
- user 'interpretation process' research
- follow-up research into reported instances of qualification misuse

Reflections on validation

- Validation isn't a one-off, post hoc project; it's an **ongoing programme**
- A lot of validation evidence/analysis arises as a **natural by-product** of designing, developing and delivering a qualification
- Often, validation research targets a **particular step** ('micro-validation')
 - investigates the various features and processes which operationalise that step, to test whether they have been effectively designed
 - can you **describe** the features and processes which operationalise each step?
 - can you **explain** how each one helps to ensure the overall validity of the qualification?
 - can you produce any **evidence/analysis** to back-up that explanation?
- Sometimes, validation research targets the **overarching claim** ('macro-validation')
 - investigates other sources of evidence/analysis, to test the claim that it is possible to measure the target proficiency by implementing the assessment procedure