Unmasking Equating Challenges in Licensure and Certification

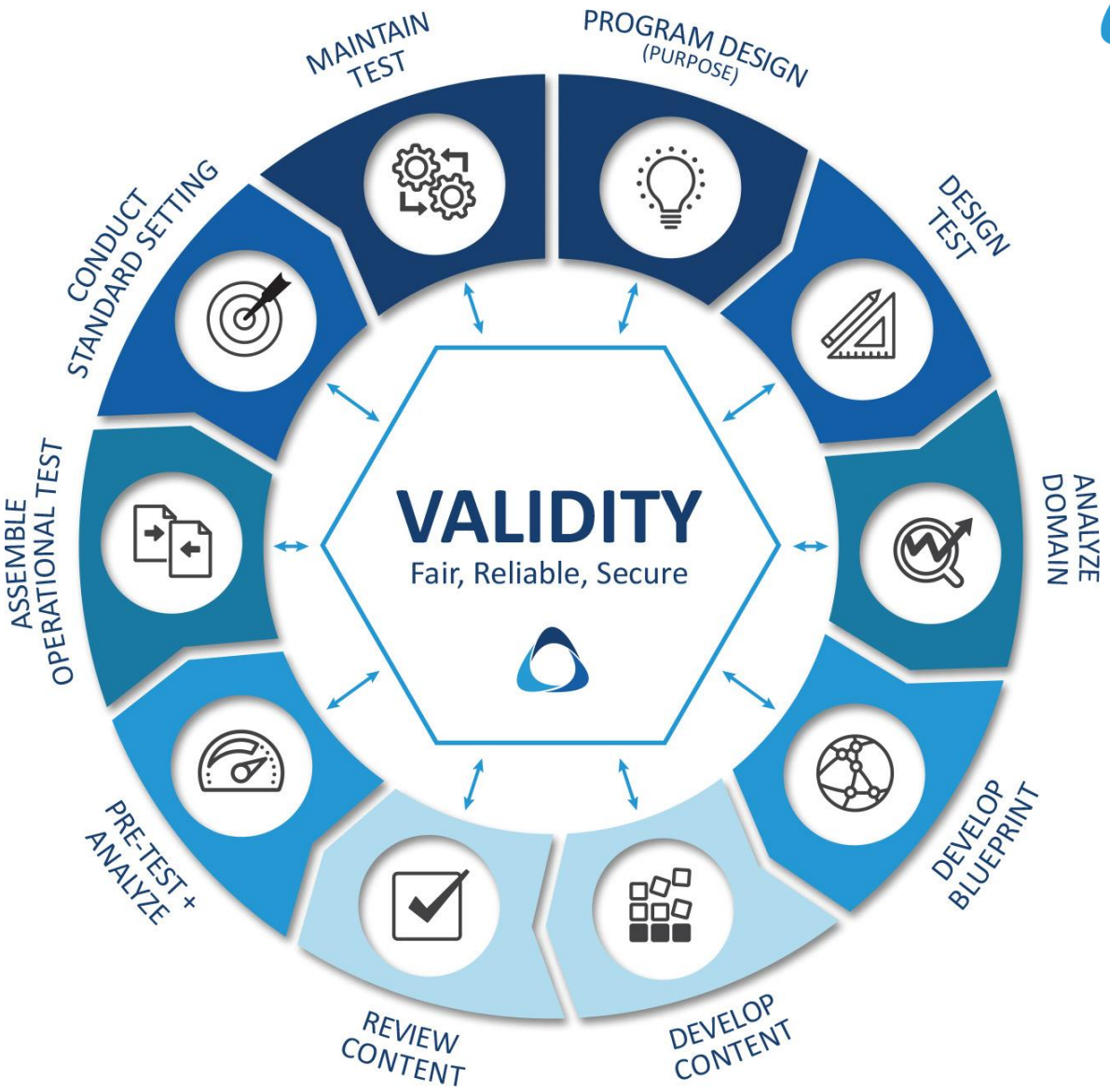
Amanda A. Wolkowitz, Ph.D.
Senior Psychometrician
Alpine Testing Solutions, Inc.

WHAT EQUATING CHALLENGES ARE UNIQUE TO LICENSURE AND CERTIFICATION?

- Practical programmatic constraints and requirements
- Focus on pass/fail decisions, which are often needed for a promotion or job
- Representativeness of sample, sample size, and making exams available to all candidates during the pandemic
- Licensure and certification exams serve the purpose of protecting the health, safety, and welfare of the public; comparability of results on parallel forms of the same exam is critical

VALIDITY WHEEL

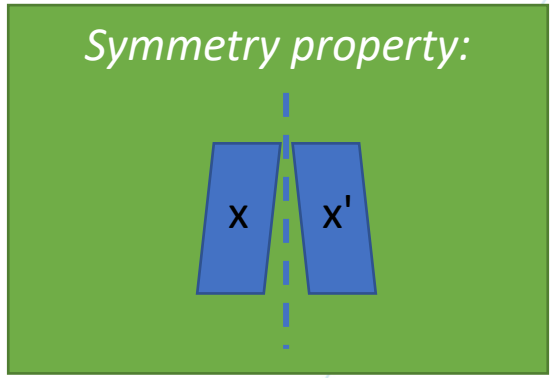
Any change in the exam may impact the validity of the score interpretation.



FACTORS THAT MAY SIGNIFICANTLY IMPACT EQUATING RESULTS

Factors that may impact equating:

- Blueprint updates
- Test specification changes
- New test administration methods



TAKING A STEP BACK: WHAT HAPPENED IN 2020 THAT MAY IMPACT EQUATING?

- Businesses and schools closed and/or went virtual causing a change in how learning occurred
- Testing centers closed, then opened with limited capacity causing candidates to put jobs on hold and (re)prepare to take an exam at a later time
- Certification and licensure programs made adjustments to deadlines and explored remote proctoring (RP) options, leading to redesigns of exams and process changes
- Many candidates tested later than they originally planned and potentially in a different testing environment causing potential angst and uncertainty for candidates
- If testing at a TC, masks are now required
- Need for licensed and certified professionals did *not* decline

HOW MAY THESE CHANGES IMPACT EQUATING?

- Proctoring methods
- Test specifications
- Comparability study
- Equating plan
- Additional documentation

PROCTORING METHODS

Proctoring method is not an equating issue, but the data used for equating is in question if it contains construct-irrelevant variance that leads to incomparable scores on the same or parallel forms.

PROCTORING METHODS

How do you remove construct-irrelevant variance?

- In assessing comparability of scores from different proctoring methods, all confounding variables need to be accounted for so that the proctoring method is the only variable compared. Otherwise, the equating results may be biased (Shadish, Cook, & Campbell, 2002).
- Identifying confounding variables is a challenging task for licensure and certification programs given the limited amount of demographic variables available.
- Few studies currently exist that compare RP to TC

PROCTORING METHODS

How do we maintain accreditation if we start equating with RP data?

- Accreditation organizations have begun accepting RP if there is evidence that the deliveries of the exam administrations are equivalent, consistent, and secure (e.g., ISO 17024 – Standard 9.3.1, NCCA – Standard 18).
- **DOCUMENT, DOCUMENT, DOCUMENT** your data collection, any studies, and equating methodologies. You cannot over document.

TEST SPECIFICATIONS

- Document with empirical evidence how changes to parallel forms of an exam maintain the validity of the score interpretation.
- Assess the extent to which the properties of equating still hold.

TEST SPECIFICATIONS

How does a change to “forward only” testing impact the equivalence of forms?

- The impact will differ by program, but this provides a fundamentally different testing experience.
- To assess impact, complete a comparability study by accounting for confounding variables that contribute to construct-irrelevant variance.
- If there truly is a difference in performance, to what extent is the difference attributable to item difficulty vs. testing experience? Will the differences fade over time as candidates adjust to the new format?
- Studies comparing PBT and forward-only CBT did not find significant differences in performance (e.g., Eaves & Smith, 1986; Lee, Moreno, Sympson, 1986).

TEST SPECIFICATIONS

How does shortening exams impact equating?

- 11 example questions to ask when determining how shortening an exam will impact equating:
  1. To what extent did the content of the test blueprint change?
  2. Were the number of items on the exam proportionally reduced according to the blueprint?
  3. Were the statistical properties of the reduced forms comparable to the pre-Covid forms?
  4. Was the exam time appropriately adjusted?

TEST SPECIFICATIONS

How does shortening exams impact equating?

- 11 example questions to ask when determining how shortening an exam will impact equating:
  1. To what extent did the content of the test blueprint change?
  2. Were the number of items on the exam proportionally reduced according to the blueprint?
  3. Were the statistical properties of the reduced forms comparable to the pre-Covid forms?
  4. Was the exam time appropriately adjusted?
  5. Did the presentation of items remain standardized?
  6. Has the interpretation of the score changed?
  7. How much shorter is the exam compared to the original exam?
  8. Is the shortening removing redundancies or essential measurement opportunities?

TEST SPECIFICATIONS

How does shortening exams impact equating?

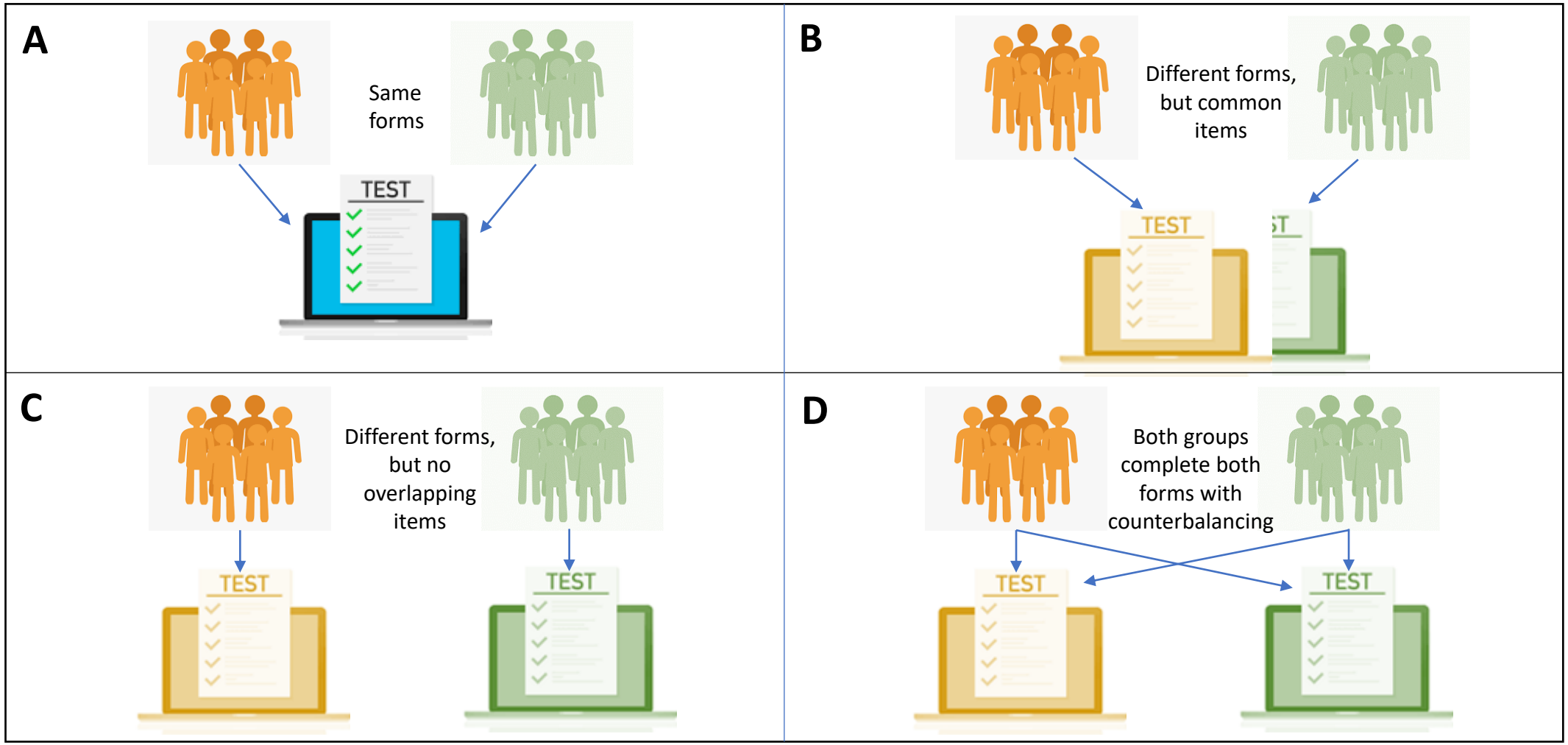
- 11 example questions to ask when determining how shortening an exam will impact equating:
  1. To what extent did the content of the test blueprint change?
  2. Were the number of items on the exam proportionally reduced according to the blueprint?
  3. Were the statistical properties of the reduced forms comparable to the pre-Covid forms?
  4. Was the exam time appropriately adjusted?
  5. Did the presentation of items remain standardized?
  6. Has the interpretation of the score changed?
  7. How much shorter is the exam compared to the original exam?
  8. Is the shortening removing redundancies or essential measurement opportunities?
  9. Is the sample completing the reduced length exam representative of the population?
  10. Were stakeholders given sufficient notice of the change?
  11. Which is more acceptable and appropriate for purposes of maintaining an equivalent interpretation of the pass/fail score: Judgement error from conducting a new standard setting study or equating error resulting from changes in the test specifications?

COMPARABILITY STUDIES

- Purpose is to provide evidence of the comparability of cut scores
- Support validity of score interpretation
- Provide evidence of fairness
- To conduct a study:
  - Minimize construct-irrelevant variance by accounting for confounding variables in studies
  - Consider both sample size and error in studies

COMPARABILITY STUDY FOR DIFFERENT PROCTORING METHODS

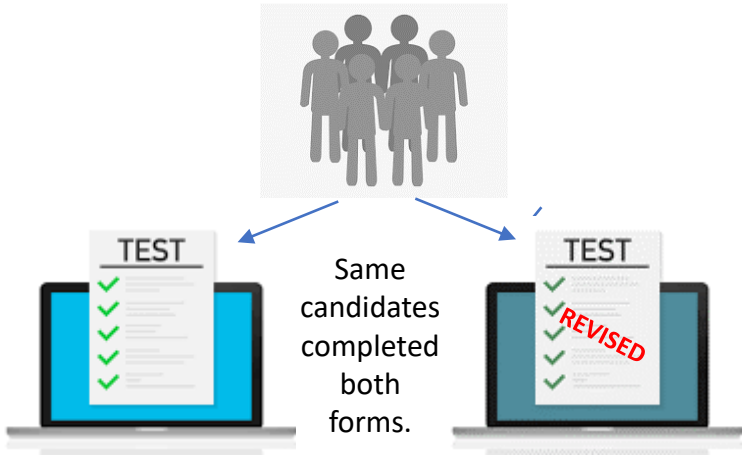
How do you design a comparability study to determine if the proctoring method impacts equating?



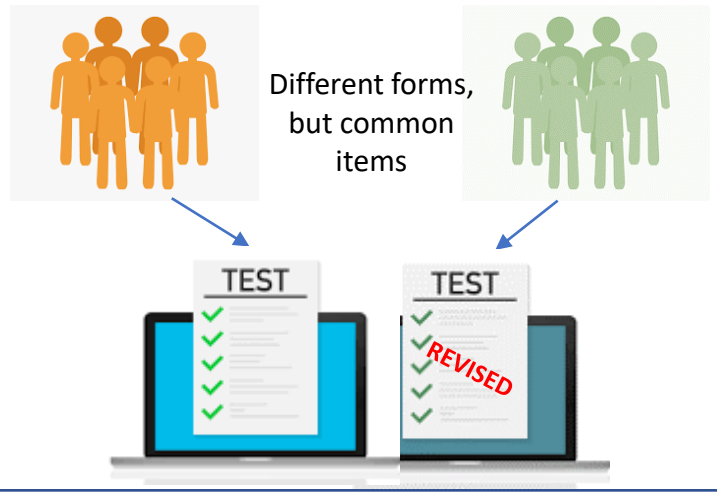
COMPARABILITY STUDY FOR TEST SPECIFICATION CHANGES

How do you design a comparability study to determine if changes to the test specifications impact equating?

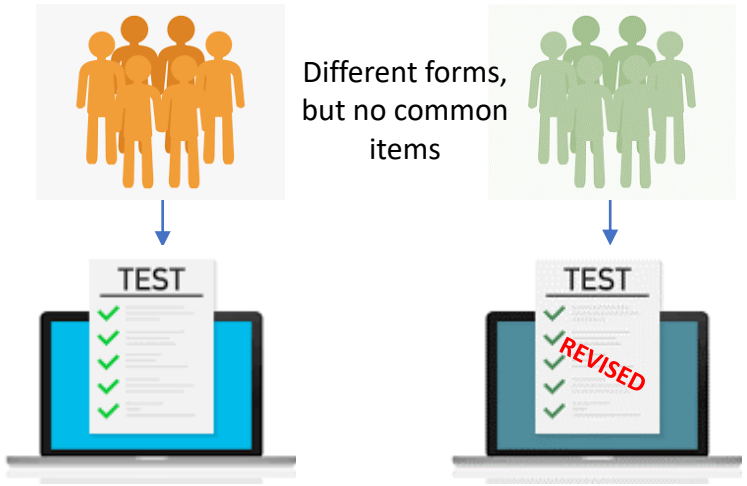
A



B



C



- *Compare historical item performance*
- *Compare pass rates*
- *Document plan, rationale, analysis, findings, threats to equivalence, and how threats were addressed.*

Pre-Pandemic

Pandemic

EQUATING PLAN

- Document the plan to establish equivalence
- Consider stability, representativeness, and statistical characteristics of anchor items
- Always do a reasonableness check on both the plan and results

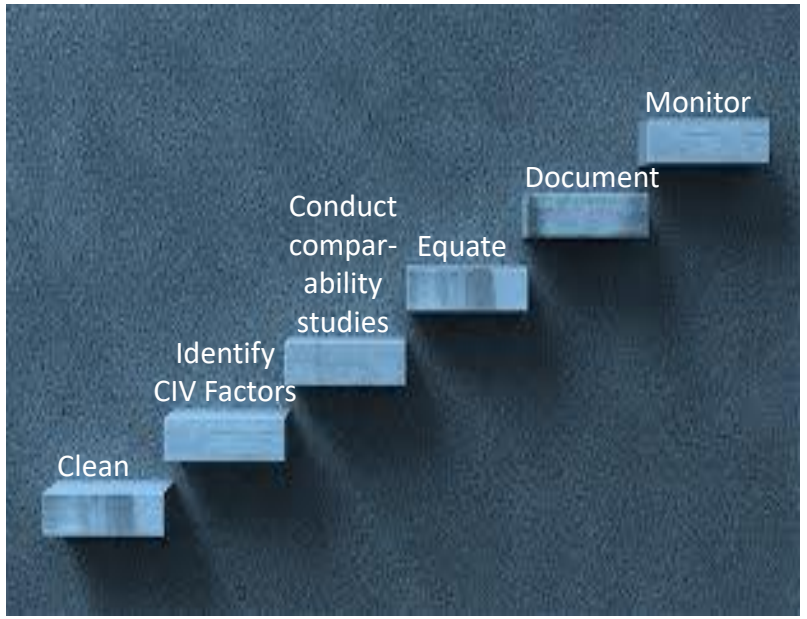
EQUATING PLAN

How does the Covid-effect impact equating?

- Small sample sizes
- Unexpected changes in performance
- Anchor block
- Questionable pretest data
- Delayed scoring
- Research
- More documentation to support the validity of the score interpretation

EQUATING PLAN

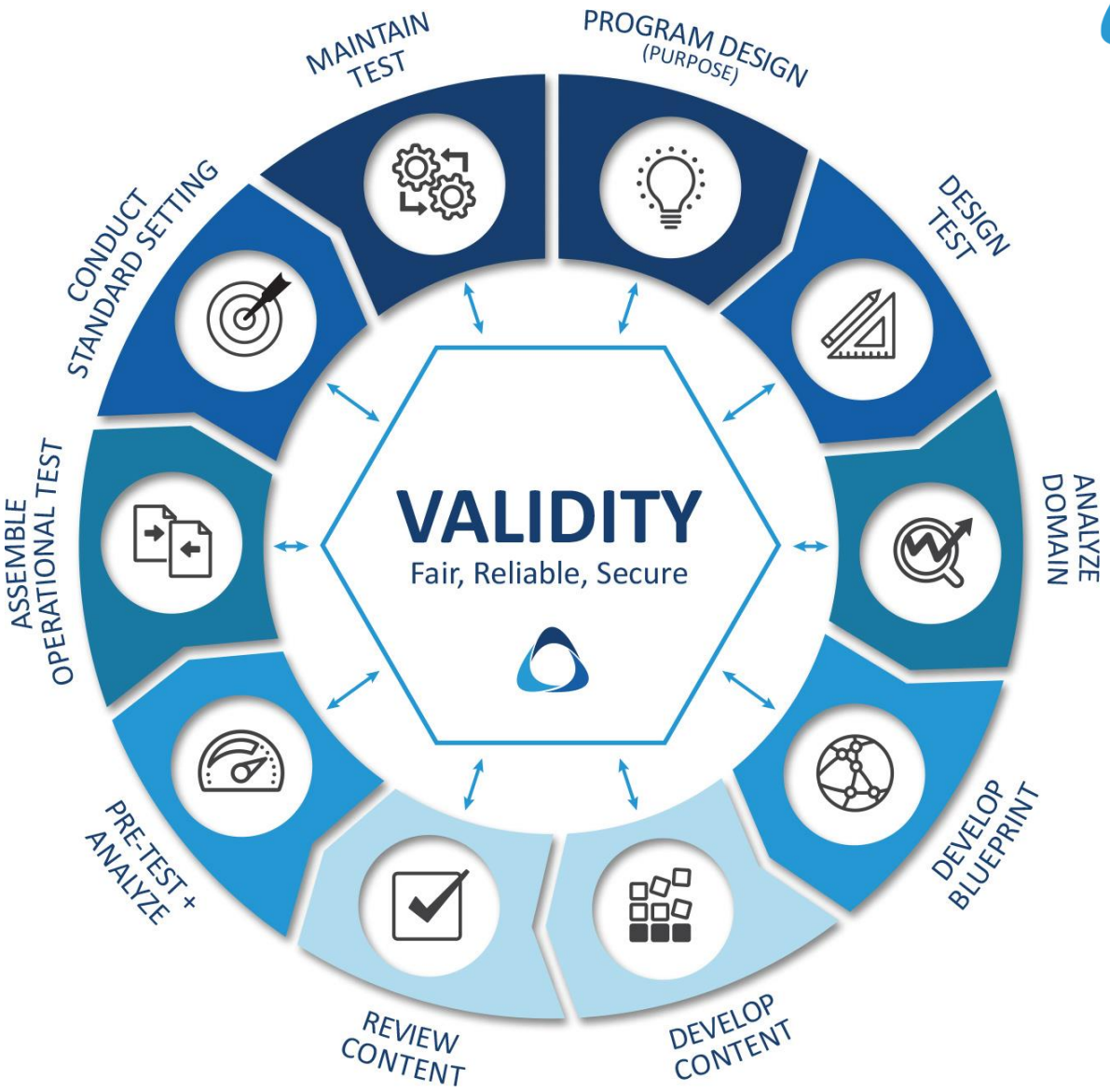
- CLEAN the data of flagged or invalid attempts
- IDENTIFY FACTORS IMPACTING CONSTRUCT-IRRELEVANT VARIANCE (CIV)
- CONDUCT COMPARABILITY STUDIES
- EQUATE and do reasonableness checks.
- DOCUMENT equating plans, decisions, rationales, results, threats, and mitigations.
- MONITOR data during future administrations and continue to investigate any differences.



VALIDITY WHEEL

The goal of licensure and certification programs remains unchanged: to protect the health, safety, and welfare of the public.

Programs should take appropriate actions to ensure validity of the score interpretation remains consistent through any changes due to the pandemic or otherwise.



Alpine Testing Solutions

Thank you!