An Introduction to “Shirts, Hurdles, and Tunnels”

Or how to:

1. Diversify Your STANAG 6001 tests with Computer Adaptive Tests.
2. Assign STANAG Levels to Test Scores.

BILC Seminar, 16 October 2014

Ellwangen, Germany

Ray Clifford

Attending BILC is always a learning experience.

- Much of that learning takes place outside of the seminar room.

- So here is a test of your aptitude for incidental learning…

What are these phenomenon?

By Dinkelsbühl…

Near our hotel…

Answer: They are labyrinths.

- But what is their purpose?
- Hint: People have been seen walking around in these circles – but WHY?

Labyrinth

- People were seen walking around in these circles – but WHY?
  - Explanation #1:

    They are investigating “crop circles” left by alien space ships.

Labyrinth

- People were seen walking around in these circles – but WHY?
  - Explanation #1:

    They are investigating “crop circles” left by alien space ships.
  - Explanation #2:

    They have found that in times of stress, going around in circles helps them unwind.

Labyrinth

- People were seen walking around in these circles – but WHY?
  - Explanation #1:

    They are investigating “crop circles” left by alien space ships.
  - Explanation #2:

    They have found that in times of stress, going around in circles helps them unwind.
  - Explanation #3:

    They are innovators, and it is obvious that people who go around in circles are doing something revolutionary.

And did you notice these similarities?

- We are
  - Staying in a religious retreat.
  - Attending a military conference.
  - Discussing language learning.
- This situation is not new. The first language test:
  - Is recorded in a religious record.
  - Was developed by the military.
  - Recognized the difficulty of language learning.

The “First” Language Test

- Reference: Judges, Chapter 12.
- The military situation.
  - The Gileadite army had won a battle against the Ephraimites at the “passages of Jordan”.
  - The battle progressed so rapidly that many Ephraimite soldiers were left behind the advancing Gileadite forces.
  - The stranded Ephraimite soldiers then had to cross through the Gileadite lines to rejoin their own army.

A “Pass / Fail” Language Test (Judges 12: 5-6)

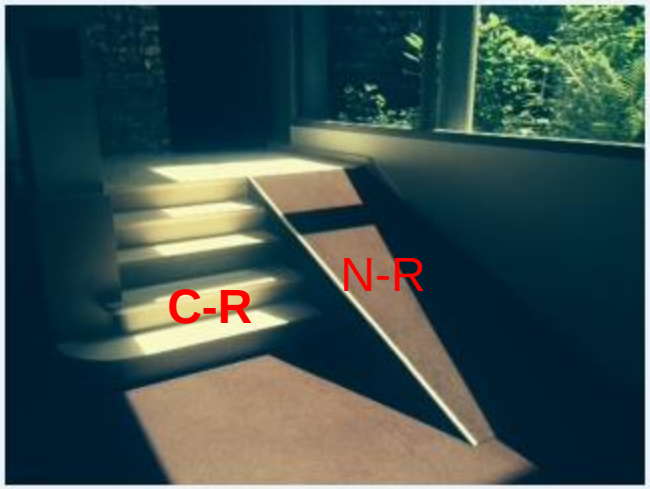
5. …when those Ephraimites which were escaped said, *Let me go over*, that the men of Gilead said unto him, *Art thou an Ephraimite?* If he said, *Nay*;
6. Then said they unto him, *Say now Shibboleth*; and he said *Sibboleth*; for he could not frame to pronounce it right. Then they took him, and slew him…and there fell at that time of the Ephraimites forty and two thousand.

Ramps and Steps

- What does this feature of the hotel in Brugges have to do with STANAG 6001 testing?

These ramps and steps

are metaphors for two different approaches to test design and scoring.



- The **steps** represent a Criterion-Referenced (C-R) test design.
- The **ramp** represents a Norm-Referenced (N-R) test design.

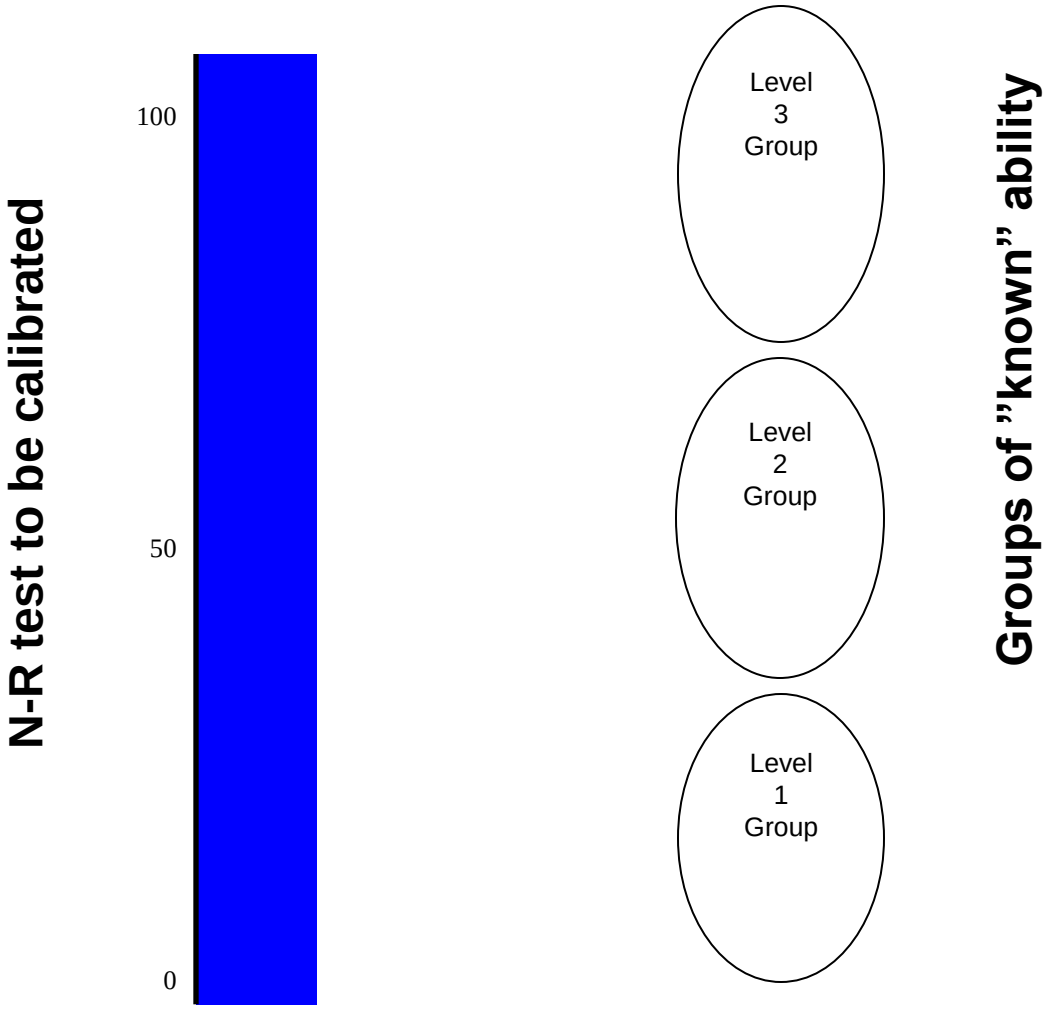
STANAG 6001 Tests: N-R or C-R?

- Every base STANAG 6001 level description is a step, defined by 3 components:
  - **Context** (Content/Topics)
  - Communication **Tasks**/Functions.
  - **Accuracy** (and precision) expectations.
- At each step (base level), the TCA elements differ from those at other steps.
- Thus the levels are not really a “scale”, but a hierarchy of Criterion-Referenced ability levels or steps.

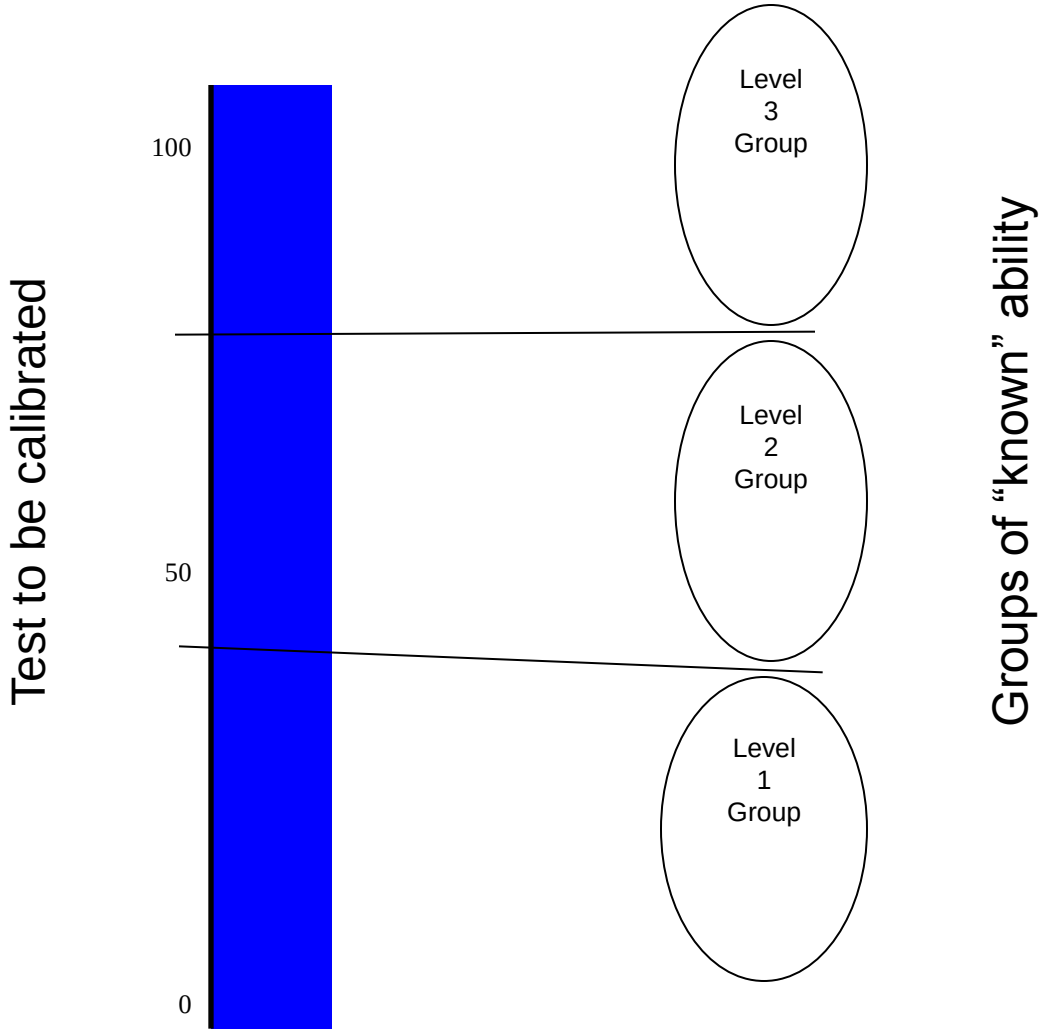
Still, most tests use a N-R (or “Ramp”) Design

- These N-R tests generate a single, compensatory score.
- That score is converted using “standard setting” judgments that estimate the ramp score’s equivalent “step height”.
- The conversion process is complicated by the fact that the single “ramp” score is essentially a total or average score attained across all of the tested levels.

A Traditional Method Of Setting Cut Scores

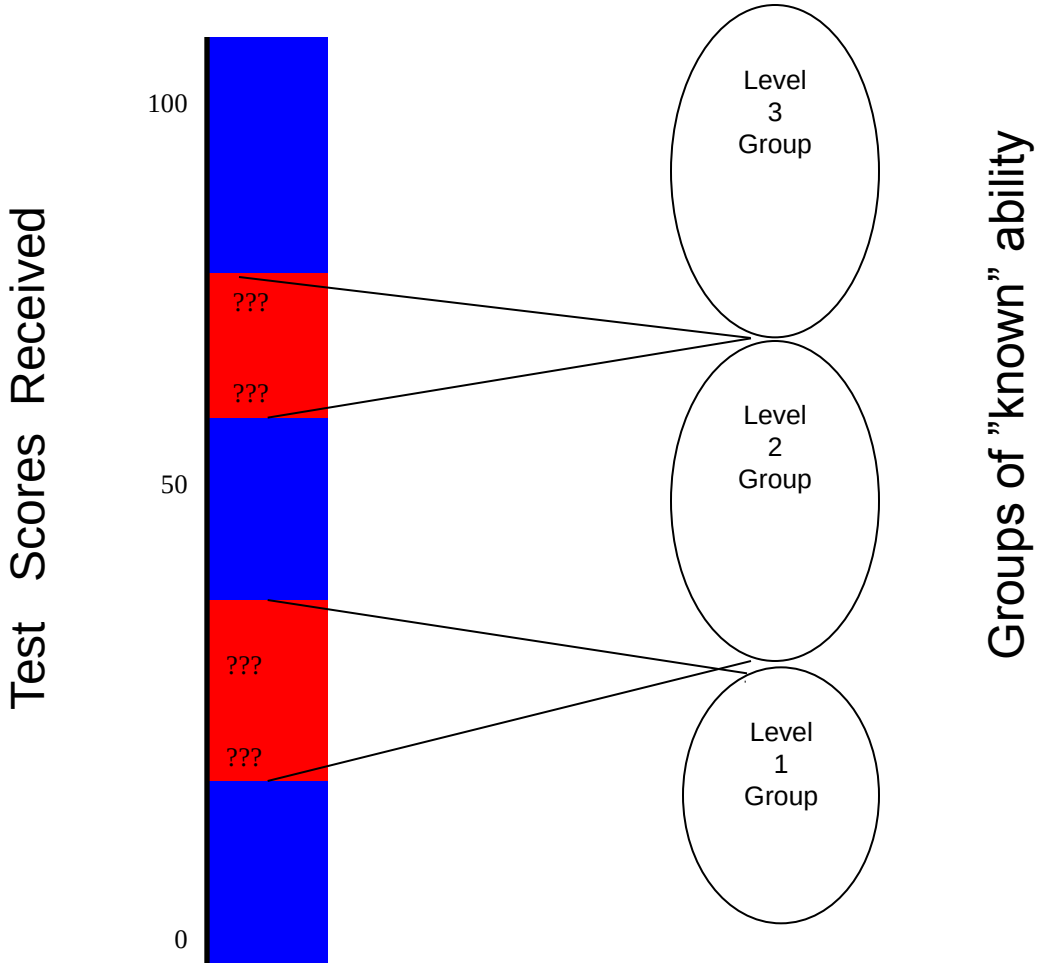


The Results One Hopes For:



The Results One Always Gets

(Some test takers score below and some score above their “known” ability.)



No matter where the cut scores are set, they are wrong for many test takers.

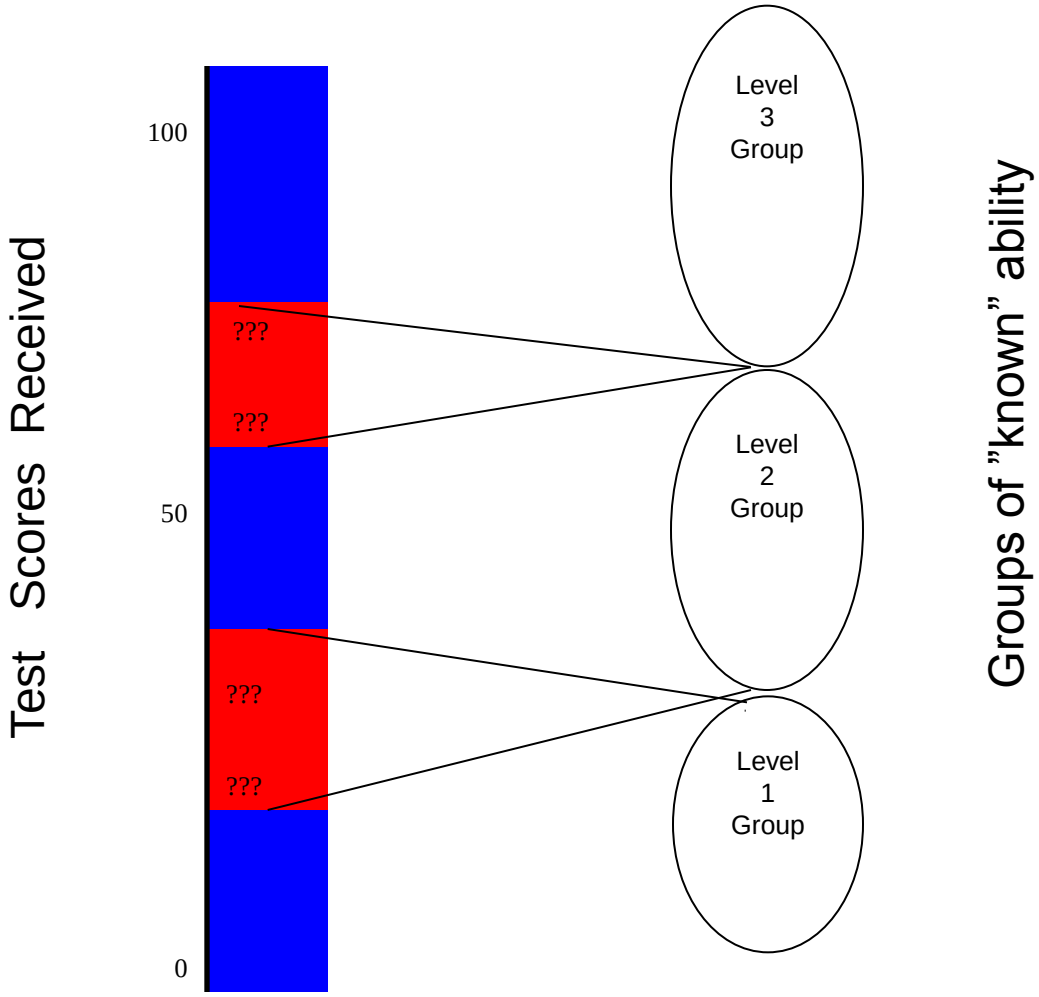


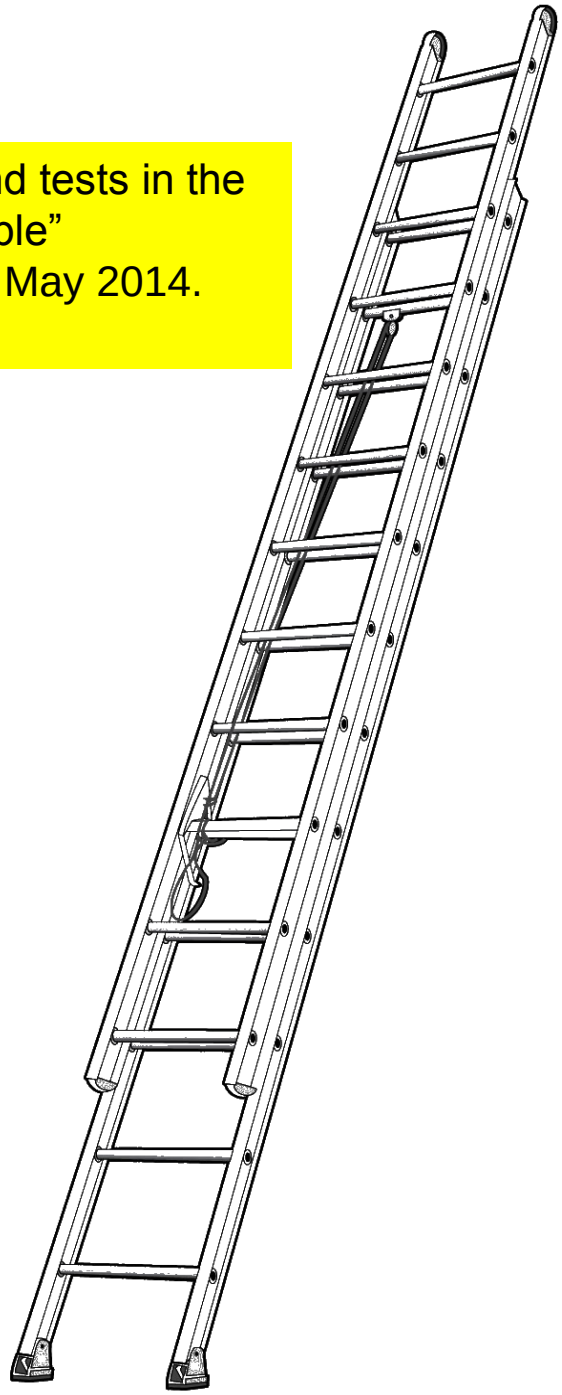
Figure 1. Traditional N-R Tests.

Adapted from a presentation by Glenn Fulcher “Standards and tests in the military domain: The arbitrary, the absolute, and the achievable” delivered at the NATO BILC Conference, Brugge, Belgium. 5 May 2014. Used with permission.

One of many mistaken ideas

“…when we speak of ‘setting performance standards’ we are…referring to the…concrete activity of deriving cut points along a score scale” (Cizek and Bunch, 2007, p. 14).

Standards aren’t set by dividing the ladder into ranges; rather each standard must be fully described.



Is there a better way than indirect extrapolation to assign proficiency levels?

Would close adherence to the STANAG 6001 TCA criteria improve testing accuracy?

Language Learning Considerations

- Language learners do not completely master the communication tasks and topical domains of one proficiency level before they begin learning the skills described at the next higher level.
- Usually, learners will have developed conceptual control or even partial control over the next higher proficiency level by the time they have attained sustained, consistent control over the lower level.

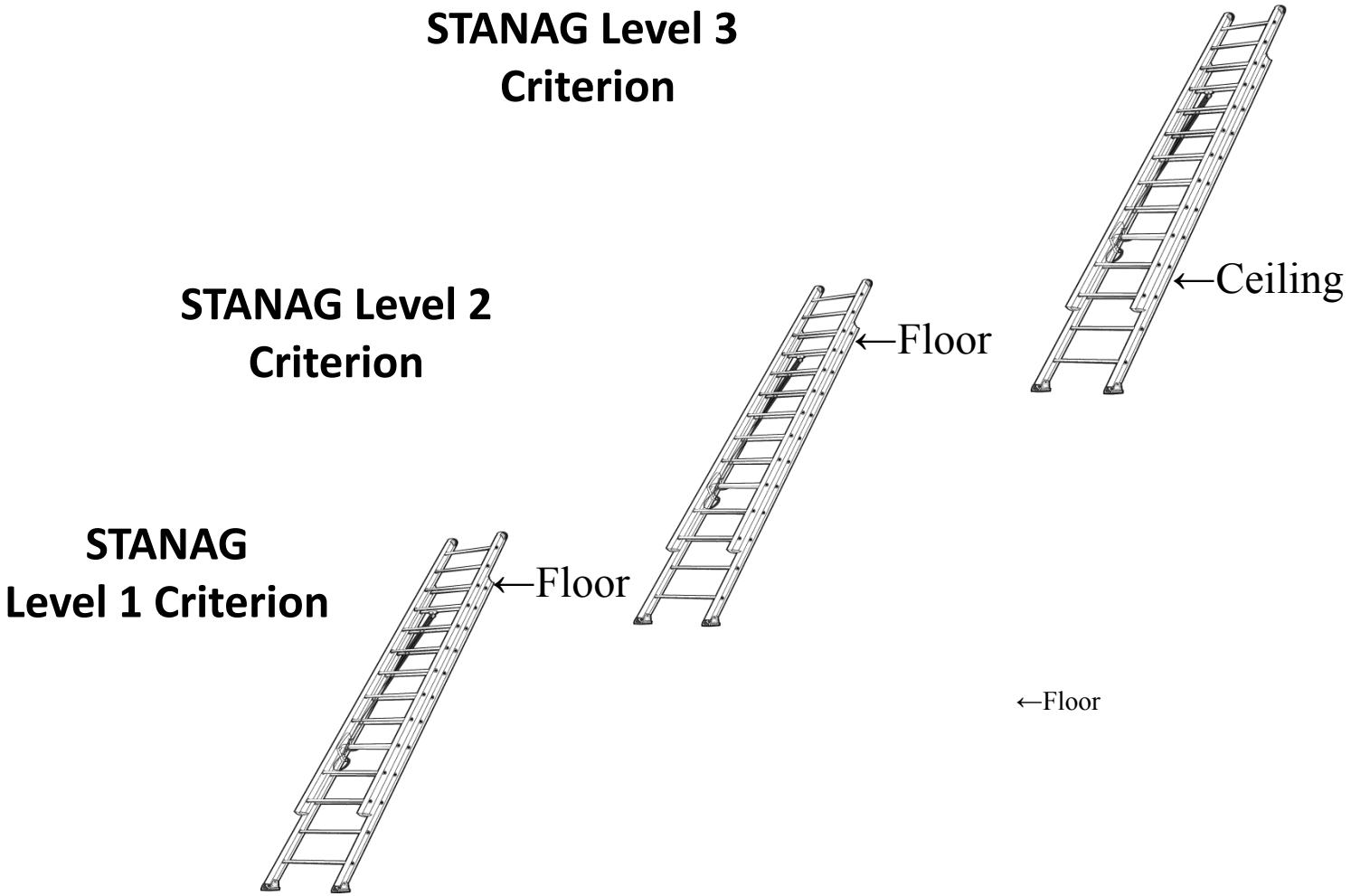
Using C-R Scoring

- Calculate the person's ability score for each step.

  (Only the error variance for one step is included in each of those scores.)

- Use con-compensatory scoring to identify each person's "floor and ceiling" ability.
  - The floor is the highest level where mastery of the TCA criterion is demonstrated.
  - The ceiling is the first level where the person's performance does not meet all the criterion.

Figure 2. A C-R Proficiency Test Design



Summary

- Why are “Floor” and “Ceiling” ratings used?
  - Criterion-Referenced testing requires a separate score for each criterion.
  - Therefore, C-R testing uses non-compensatory, level-specific scoring.
  - These independent scores explain ability distinctions that would be regarded as error variance in multi-level tests that report only a total test score.

And now lets talk about hurdles and shirt sizes…..