Physics Education Research (PER) seminar: all about concept inventories

Michael L Allen
mlfa.astro@gmail.com

2015-11-10

Goals for today's seminar

1. define and/or describe concept inventories
2. describe how they are constructed
3. apply them in a course setting
4. interpret results using item response theory

What are concept inventories?

A concept inventory is a test designed to quantify a person's knowledge/skills/attitudes in a narrowly-defined discipline.

1. target audience is usually college & professional students
2. usually multiple choice
   - exactly one correct response (exception: attitudes)
   - incorrect responses, called distractors, are constructed from research results
3. used to measure degree of mastery; there is no passing or failing grade
4. usually proprietary, i.e., not freely available, users have to agree to conditions of use (e.g., fee, secrecy, data-sharing)

What are some examples of concept inventories?

- Knowledge
  - ADT: astronomy diagnostic test (CAER 2002)
  - BEMA: basic electricity & magnetism assessment (Ding et al 2006)
  - FCI: force concept inventory (Hestenes 1992)
  - LPCI: lunar phases concept inventory (Lindell & Olse 2004)
  - LSCI: light & spectroscopy concept inventory (Bardar et al 2007)
- Skills
  - CCTST: California critical thinking skills test (Facione 1990)
  - LASSI: learning and study strategies inventory (Weinstein et al 2016)
  - VARK: visual aural reading kinesthetic (Fleming 2011)
- Attitudes
  - SATA: survey of attitudes toward astronomy (Zeilik)

How are concept inventories constructed?

1. choose an area of interest
2. interview students before, during, and after instruction
3. record student responses and map them to a common concept lexicon
4. construct multiple choice questions
   - the distractors (incorrect choices) are modeled after student responses
5. validate the questions
   - by experts

Examples of misconceptions: Moon phases

From student interviews, misconceptions can be recorded concerning the Moon. A wide variety of descriptions is mapped to a few common ideas:

- Earth's shadow causes Moon phases
- clouds cause Moon phases
- Earth's rotation causes Moon phases
- Moon's rotation causes Moon phases
- the Moon does not rotate
- Moon's dark side sometimes faces Earth
- the Moon is self-luminous
- the Moon is viewable only at night
- the Moon completes one orbit per day

A multiple choice question can be constructed and validated from this data.

Examples of validation

A validated question is one for which a group of experts agree upon what the question is asking.

Let's invalidate a few...

Examples of validation cont'd

Is anything wrong with this question?

1. Moon phases are _____.
   a caused by Earth's shadow
   b important for time keeping
   c the changing appearance of the Moon
   d pretty to look at

Examples of validation cont'd

Is anything wrong with this question?

1. Which is the most important Moon phase?
   a New
   b First quarter
   c Full
   d Third quarter

Examples of validation cont'd

Is anything wrong with this question?

1. What is the cause of Moon phases?
   - a clouds
   - b the Moon's dark side
   - c the changing aspect of the Moon is an effect of our forced perspective from the surface of the Earth, where we view only the facing hemisphere of the Moon, some of which is illuminated by the Sun depending upon where the Moon is in its orbit with respect to the Sun and Earth

Is anything wrong with this question?

1. What is NOT the cause of Moon phases?
   a clouds
   b Moon's dark side
   c viewing perspective
   d both (a) and (b) but not (c)
   e both (a) and (c) but not (b)
   f Both (b) and (c) but not (a)
   g none of the above

Examples of validation cont'd

Is anything wrong with this question?

1. What is the cause of Moon phases?
   a budget cuts
   b nude beaches
   c the Illuminati
   d viewing perspective

Examples of validation cont'd

Take home lesson: validation is a complex process involving content, format, word usage, and language conventions (spelling, grammar, punctuation).

How to use concept inventories in a course setting

Compare pre-instruction results with post-instruction results.

Compare post-instruction results with some baseline.

Both require use of item response theory (IRT).

Examples of quantitative measures in item response theory

1. Difficulty factor.
2. Discrimination index.
3. Normalized learning gain.

Examples of quantitative measures in item response theory cont'd

1. Difficulty factor

Difficulty factor is the fraction of students who got the answer correct.

Difficulty $= 0$ implies that no one got the correct response.

Difficulty $= 1$ implies that everyone got the correct response.

Examples of quantitative measures in item response theory cont'd

2. Discrimination index (DI)

Discrimination index is the performance of the students on an individual question, compared with the performance of the entire group.

$DI = f_{upper} - f_{lower}$

f_{upper} is the fraction of the upper quartile (27%) that correctly answered the question.

f_{lower} is the fraction of the lower quartile (27%) that correctly answered the question.

Examples of quantitative measures in item response theory cont'd

2. Discrimination index (DI)

Limits: $-1 \leq DI \leq +1$

$DI = +1$ implies that everyone who got the answer correct to this question also earned the highest overall score on the test. This question is a good discriminator between well-performing and poorly-performing students.

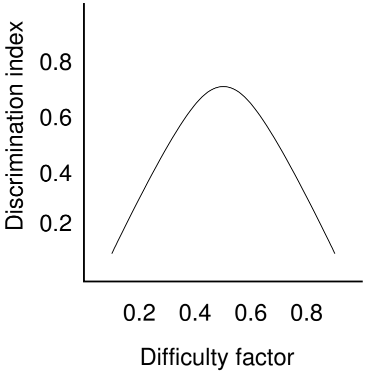
$DI = 0$ implies that those students who got the answer correct were divided evenly among those who did best overall and those who did worst overall. This question does not discriminate.

$DI = -1$ implies that those students who got the answer correct also earned the lowest overall score on the test. This question anti-correlates with students who did well overall.

Examples of quantitative measures in item response theory cont'd

1. Difficulty factor, and 2. Discrimination index

A plot of 2. versus 1. yields a cap-shaped ∩ trend where the "ideal" question is at the top of the cap.



Examples of quantitative measures in item response theory cont'd

An interesting result

For 5 consecutive semesters, I gave a multiple choice final exam in Astronomy 450: Life in the universe. I replaced 5% of the questions with the worst discrimination index (DI), and left the remaining questions the same. In each successive semester, the final exam scores showed a bi-modal distribution, and the distance between modes increased.

I now have a well-discriminating question bank.

Examples of quantitative measures in item response theory cont'd

3. Normalized learning gain

$Gain = \frac{post - pre}{100\% - pre}$

$-1 \leq Gain \leq +1$

Examples of quantitative measures in item response theory cont'd

More complex measures of validity exist, e.g., Cronbach alpha, Kuder Richardson 20, and contingency tables.

Results from intro astronomy courses

ADTv2 (21 items): Astronomy 150, spring 2015, N=60

Top 5 learning gains:

Q	PRE	POST	GAIN	SUBJECT
16	0.62	0.96	0.89	Center of the universe
12	0.47	0.77	0.57	Solar system to stars distance scale
17	0.30	0.65	0.50	Color and temperature (Wien's law)
13	0.60	0.77	0.43	Solar system distance scale
5	0.07	0.31	0.26	Speed of light

What do these have in common?
Active learning, revisited frequently.

Results from intro astronomy courses

ADTv2 (21 items): Astronomy 150, spring 2015, N=60

Bottom 5 learning gains:

11	0.62	0.54	-0.21	Earth-Moon-low orbit scale
14	0.25	0.15	-0.13	Mass and gravity
20	0.45	0.44	-0.02	Distance and angle subtended
3	0.28	0.27	-0.01	Earth-Moon scale
6	0.08	0.08	0.00	Free fall orbit

What do these have in common?

Do not appear in the course.

Wrap-up

Concept inventories and item response theory are useful for identifying and quantifying areas of strength in our students' learning.

Wish list: use of a standardized instrument to measure learning gains in our BS Physics program, specifically how our graduates measure up to our published student learning outcomes (SLOs); this would involve a concerted faculty-wide effort.