

# Classical Item Analysis in the Development of a Measure of Pre-service Elementary Teacher Understanding of Mass Density



Eric A. Hagedorn, Ph.D. (Department of Physics)  
Sudheer Rakasi, B.S. (Masters Program in Electrical Engineering)

## Classical Test Theory Item Statistics

Item	N	Mean	S.D.	Rmean	Facility	Diff	Bis	P. Bis
1	84	3.29	1.75	5.03	.369	14.34	.974	.762
2	83	3.33	1.73	3.73	.627	11.71	.389 <sup>2</sup>	.304 <sup>2</sup>
3	80	3.34	1.73	5.45	.275	15.39 <sup>3</sup>	1.010 <sup>3</sup>	.755
4	84	3.24	1.73	4.89	.107 <sup>1</sup>	17.97 <sup>1</sup>	.553	.330
5	80	3.38	1.68	3.67	.800	9.63	.505	.354
6	84	3.31	1.72	4.88	.381	14.21	.909	.713
7	83	3.36	1.68	3.79	.843	8.97	.888	.586

- N – number who responded to an item
- Mean – total score mean for everyone who completed this item
- S.D. – standard deviation for everyone who completed this item
- Rmean – total raw score for everyone who got this item right
- Facility (or “p”) – proportion correct response
- Diff (or “delta statistic”) –inverse of standard normal cumulative distribution for facility, multiplied by -4 and add 13. Essentially a measure of difficulty
- Bis – biserial correlation coefficient – correlation between dichotomous variable (right “1”, wrong “0”) and continuous variable (total score)
- P. Bis – point biserial correlation coefficient – correlation between dichotomous variable (right “1”, wrong “0”) and continuous variable (total score)

$$Bis = \frac{RMean - Mean}{S.D.} \frac{facility}{h(facility)}$$

h = height of standardized normal distribution

$$P.Bis = \frac{RMean - Mean}{S.D.} \sqrt{\frac{facility}{1 - facility}}$$

## Items Ranked from Easiest to Hardest

Item	Facility	Diff
7	0.843	8.97
5	0.800	9.63
2	0.627	11.71
6	0.381	14.21
1	0.369	14.34
3	0.275	15.39
4	0.107	17.97

Even after careful item generation—deciding the content domain, identifying sub-areas, considering item types, writing items, reviewing items, etc.—an item analysis of pilot data is critical to the development of an instrument. This poster presents the item analysis of a mass density concept inventory, developed for use with preservice elementary and middle school teachers.

## Actual Items from the Understanding Density Instrument

1. Which has the greater density – 1 kg of water or 10 kg of water?
2. Which has the smallest density – 5 kg of lead or 10 kg of aluminum?
3. Which of the following blocks of iron are more dense? (shows 3 increasingly large blocks identically shaded)
4. The density of gold is 19.3 grams per cubic centimeter. What is the mass of a cube made of gold that measures 1.2 cm on a side?
  - a. 0.09 grams
  - b. 1.73 grams
  - c. 23.16 grams
  - d. 33.35 grams
5. If you place a solid object made from a material that is more dense than water into a bucket of water, what happens?
6. Which has the greater density – an entire candy bar or half a candy bar?
7. What happens if you place a solid object made from a material that is less dense than water into a bucket of water?

## What Do the Numbers Tell Us?

- Item 4 (the only item involving several mathematical calculations) was clearly the most difficult item on the instrument, but the relatively low biserial and point biserial correlation coefficients indicate that the participants who did know how to work this singularly mathematical problem did not necessarily understand density from a conceptual point of view.
- Item 2 had even lower biserial and point biserial correlation coefficients than Item 4. This implies that participants who answered this problem correctly, did not necessarily do so well on the rest of the instrument, and perhaps more tellingly, those who did not answer this item correctly probably did better on the rest of the instrument. This suggests a problem with the item. A likely one is that it requires that the participants have previous knowledge about the densities of lead and aluminum. Participants could understand density, but not know one or both of these material specific facts.
- Item 3 was the second most difficult item on the instrument, but had the highest biserial and second highest point biserial correlation coefficients. This indicates 2 positive attributes of this item: 1) participants who got it right, did well on the whole instrument, and relatedly 2) this item could discriminate between participants who did fairly well on the whole instrument from those who did very well.

## Item Response Theory (IRT) – A Computationally Heavy, but Powerful Alternative to Classical Test Theory (CTT)

The approaches presented here are from the perspective of classical test theory (CTT). While useful, additional useful insights into the effectiveness of both items and instruments can be gained using the techniques of item response theory (IRT).

