Item Analysis in the Development of a Nanoconcepts Inventory

Eric A. Hagedorn, Ph.D. (Department of Physics)
Marian Manciu, Ph.D. (Department of Physics)
Vishal Kapoor, M.S.(Doctoral Program in Environmental Science)

Classical item statistics such as facility and "diff" being based on the percentage of the sample who got a particular item correct provide insight into the relative difficulties of the items, based purely on the abilities of the participants in the sample.

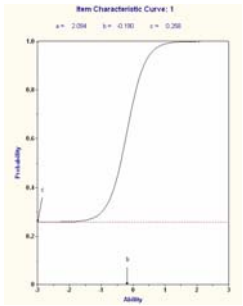
The biserial and point biserial correlation coefficients indicate how well a correct answer on a particular item correlates (or predicts) a higher total score. This allows insight into how well an item discriminates between students of varying abilities. A good item is gotten right more often by participants who have higher total grades.

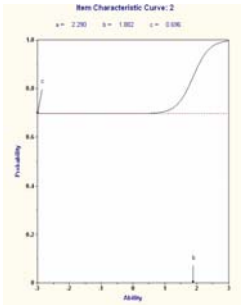
Item parameters obtained through IRT modeling not only provide information about item difficulty and discrimination, but do so in a way that is independent of the ability level of the participants in the sample. This eliminates the possibility of an item or test being considered "difficult" simply because the sample of participants who took it were of lower ability. In addition, certain IRT models take into account the effect of participant guessing on estimates of both item difficulty and discrimination.

This study presents results from an analysis of the item level responses of 205 high school students who completed the "Nanotechnology Survey" version a1. Both classical and IRT approaches were used. The classical item statistics are presented in the table below and the first table to the right. What is known as the 3 parameter model, was used to calculate IRT parameters for these same items. These are presented in the right-most table. Selected item characteristic curves are shown as well.

Item	N	Mean	SD	Rmean	Facility	Diff	Bis	P.Bis
1	205	8.63	2.97	9.45	0.678	11.15	0.522	0.401
2	205	8.63	2.97	9.07	0.707	10.82	0.302	0.228
3	205	8.63	2.97	9.65	0.483	13.17	0.413	0.330
4	205	8.63	2.97	11.14	0.180	16.65	0.579	0.396
5	204	8.61	2.96	10.44	0.167	16.87	0.412	0.276
6	205	8.63	2.97	9.72	0.556	12.44	0.515	0.409
7	205	8.63	2.97	9.56	0.483	13.17	0.376	0.300
8	205	8.63	2.97	10.65	0.322	14.85	0.611	0.469
9	205	8.63	2.97	9.85	0.517	12.83	0.531	0.424
10	205	8.63	2.97	11.21	0.093	18.30	0.485	0.278
11	205	8.63	2.97	10.16	0.312	14.96	0.453	0.346
12	205	8.63	2.97	8.97	0.839	9.04	0.383	0.255
13	205	8.63	2.97	9.12	0.254	15.65	0.128	0.095
14	205	8.63	2.97	10.75	0.215	16.16	0.524	0.373
15	205	8.63	2.97	12.06	0.078	18.67	0.618	0.336
16	204	8.65	2.96	9.34	0.725	10.60	0.509	0.380
17	202	8.68	2.96	9.30	0.644	11.53	0.362	0.282
18	201	8.68	2.97	9.17	0.144	17.25	0.105	0.068
19	200	8.70	2.96	9.10	0.155	17.06	0.086	0.057
20	200	8.70	2.96	9.66	0.350	14.54	0.304	0.236
21	198	8.72	2.97	10.00	0.227	15.99	0.326	0.234
22	197	8.73	2.97	9.85	0.173	16.78	0.256	0.173
23	197	8.73	2.97	10.13	0.391	14.11	0.481	0.379

The biserial and point biserial correlation coefficients highlighted in red are small enough (less than .20) that they suggest a problem with the items they represent. Item 13 asks which shape of four possibilities would maximize the surface area of a piece of clay. Item 18 asks which of four objects is similar in size to a red blood cell. Item 19 asks for the approximate number of molecules of water in a raindrop. Participants who answered these items correctly did not tend to do well on the other items.

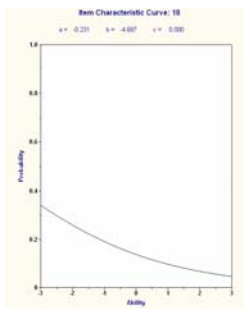


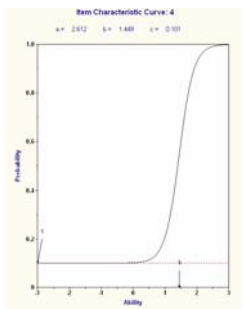


Item characteristic curves plot the probability of participants correctly answering an item as a function of ability. ICC's are typically sigmoid in shape. Participants of higher ability have a higher probability of giving a correct response. Participants of lower ability have a lower probability of giving a correct response. The item parameter, "b" is located on the ability scale corresponding to the point on the curve that indicates a 50-50 probability of success. A more difficult item has a higher b value, while an easier one has a lower value. The item parameter, "a" is proportional to the slope of the ICC at "b." The higher the value of "a," the steeper the slope, and the more the item discriminates between participants who are close in ability. The "c" parameter is known as the "pseudo-chance level" parameter and estimates the probability of lower ability participants guessing the correct response.

The "b" parameter for Item 1 above (-.190) suggests that this item was fairly easy: an individual with slightly lower than average ability has a 50-50 chance of choosing the correct response. The "a" parameter of 2.094 indicates that this item discriminates well between participants above and below this ability level. The "c" parameter of .258 is slightly larger than the probability of accurately guessing by randomly picking 1 out of the 5 available choices. Item 1 deals with atomic representations of crystalline solids.

Item 2 above is markedly different than Item 1. It is much more difficult but has the unusual characteristic that low ability guessers have roughly a 70% chance of guessing correctly! Item 2 asks why neurons have such highly branched axons and dendrites.





Item 18, which was identified as problematic using classical analysis, has an item characteristic curve indicating that higher ability participants have less of a chance of answering this item correctly than do lower ability ones! In fact, no one has greater than a 35% probability of answering this item correctly. This unusual behavior, as well as contradictory interpretations of the classical and IRT difficulties could also suggest a sample size too small for a proper IRT estimate of the parameters.

Item	Facility	Diff
12	0.839	9.04
16	0.725	10.60
2	0.707	10.82
1	0.678	11.15
17	0.644	11.53
6	0.556	12.44
9	0.517	12.83
3	0.483	13.17
7	0.483	13.17
23	0.391	14.11
20	0.350	14.54
8	0.322	14.85
11	0.312	14.96
13	0.254	15.65
21	0.227	15.99
14	0.215	16.16
4	0.180	16.65
22	0.173	16.78
5	0.167	16.87
19	0.155	17.06
18	0.144	17.25
10	0.093	18.30
15	0.078	18.67

Easiest ↓ Hardest

Item	a discrim	c guessing	b difficulty
19	-0.08	0.00	-12.14
18	-0.23	0.00	-4.69
12	0.25	0.00	-4.07
13	-0.20	0.06	-4.07
1	2.09	0.26	-0.19
9	0.47	0.00	-0.10
7	0.33	0.00	0.12
16	1.21	0.55	0.37
6	0.59	0.28	0.59
8	1.04	0.00	0.64
17	0.69	0.50	1.01
3	1.47	0.36	1.08
23	4.05	0.33	1.38
4	2.61	0.10	1.45
11	0.31	0.00	1.59
14	1.24	0.13	1.69
2	2.29	0.70	1.88
15	0.81	0.00	2.29
5	0.89	0.12	2.50
22	1.53	0.16	2.50
20	0.11	0.00	3.34
10	0.34	0.00	4.19
21	0.13	0.00	5.66

Easiest ↓ Hardest

Sorting item difficulties is one way to begin collectively interpreting the items on an instrument. The first table above provides both the classical item facility (percentage of participants choosing the correct response) and difficulty measure. The second table of IRT item parameters is also sorted by difficulty, but includes the discrimination and pseudo-chance parameters. Low discrimination parameters and high pseudo-chance parameters are highlighted. Items 13, 18, and 19, which had low classical discrimination indices (biserial and point biserial correlation coefficients) actually have negative "a" parameters.

Both the classical and IRT analyses in this study suggest a closer look at certain items in the "Nanotechnology Survey" version a1. The evidence from the classical analysis is fairly strong, however, because of the small sample size, conclusions based on the IRT analysis would be premature. In addition to collecting more item level data, so as to increase the sample size and re-evaluate the IRT model, participants from distinct ability groups (e.g. from each of the four quartiles) should be asked to describe their reactions to certain items as well as their thought processes in trying to answer them.