DOCUMENT RESUME

ED 352 379 TM 019 276

AUTHOR Brandon, E. P.
TITLE A Note on the Format of Ennis' Multiple-Choice Tests of Deductive Reasoning Competence.
PUB DATE Oct 92
NOTE 8p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; *Cognitive Tests; Comparative Testing; Competence; *Deduction; Foreign Countries; *Multiple Choice Tests; *Test Format; Test Items; *Thinking Skills
IDENTIFIERS *Ennis (Robert); *High Stakes Tests; Jamaica

ABSTRACT

In his pioneer investigations of deductive logical reasoning competence, R. H. Ennis (R. H. Ennis and D. H. Paulus, 1965) used a multiple-choice format in which the premises are given, and it is asked whether the conclusion would then be true. In the adaptation of his work for use in Jamaica, the three possible answers were stated as "yes" (it must be true), "no" (it can not be true based on what you are told), and "maybe" (it may be true or false). In the original investigations, the results were of no consequence for those tested, but in Jamaica the questions have been part of examinations of importance to the subjects. For this reason, the possible effect of question format was investigated by replacing "maybe" with "not necessarily." Results with 537 subjects in 1990 (using "maybe" test answers) and 474 subjects in 1991 (using "not necessarily" test answers) indicate that the change of format makes no difference to 18 valid items, but results in significant differences in response pattern for 13 of 18 invalid items, with more correct responses in most cases. Three tables present study data. (SLD)

A Note on the Format of Ennis' Multiple-Choice Tests of Deductive Reasoning Competence

E.P. Brandon
Office of Academic Affairs
UWI, Cave Hill
Barbados

The central notion of deductive logic is that of a valid argument, an argument in which the conclusion really does follow from the premisses. The standard informal explanation of validity is that it should be impossible for the premisses of the argument to be true and the conclusion false.

Investigation of people's actual competence in matters of deductive logic could use questions of the form "Does *r* follow from *p* and *q*?" But given the standard informal explanation, one could avoid any uncertainties people might feel about what it is for one statement to follow from another, by framing the question in terms of the notions of truth and falsehood. So, for instance, in his pioneer investigations of deductive logical reasoning competence, Robert Ennis (Ennis and Paulus, 1965) employed the following question structure:

Suppose you know that *p*, *q*, ...
Then would it be true that *r*?

In an adaptation of Ennis' work for use in Jamaica, the three possible answers offered (Yes; No; Maybe) are glossed as "Yes" means "It must be true, given what you are told"; "No" means "It can't be true, given what you are told"; and "Maybe" means "It may be true or it may be false. You haven't been told enough to be certain whether it is Yes or No."

Given Ennis' question format, and the standard construal of validity, when the sentences constitute a deductively valid argument the correct answer is either "Yes" or "No"; when they do not make up a valid argument the correct answer is "Maybe."

In the course of the Jamaican investigations (reported in Nolan and Brandon, 1986, and Brandon, 1990) a doubt arose concerning the question format. While the correct answers in the case of valid arguments seem conversationally appropriate, this does not seem so obviously the case for the invalid ones. Contrast these two dialogues:

(*i*) Suppose some vegetarians drink milk; would some people who drink milk be vegetarians?
Yes.

(*ii*) Suppose most teachers are women; would it be true that most women are teachers?
Maybe.

Speaking personally, my usual response in cases such as (*ii*) would be to say "No," or more fully "No, not necessarily."

In Ennis' original investigation, the results were of no consequence for those tested; subjects had no time limit to complete the questions; and they were reminded of the meaning of the answers on every page of the question booklet. The collection of most of the data in Jamaica has involved three serious departures from this set-up: the questions have formed part of an entrance examination for the Faculty of Education; there has been a time-limit for the examination; and instructions were given only on the first page of the appropriate section of the booklet. With such an increase in pressure, it is likely that interference from linguistically odd constructions would be more serious than in Ennis' original investigations.

It was decided to investigate the possible effect of the question format by using the same items in two successive years with one small change in question format: to replace "Maybe" with "Not necessarily."

The item analyses in Tables 1 and 2 report the main findings: in general the change of format makes no difference to performance on the valid items (Table 1) but a considerable difference to performance on the invalid questions (Table 2). The tables give the numbers offering each of the three possible answers or omitting to answer the question, the correct answer in each case, the difficulty index for the item (really a facility index), and finally the chi-square value and its probability for the comparison of the distribution for the two years.

Table 1 shows that in only one of the 18 valid items in the test was there a significantly different distribution of answers between the two years, though it made no difference to the difficulty of the item. In general the three forms of reasoning were of similar difficulty for the two groups.

Comparison with Table 2 shows that the overall facility of two of the three invalid types of item increased markedly and that in 13 of the 18 invalid items the distribution of answers was significantly different.

These results seem to show that the change from "Maybe" to "Not necessarily" has a marked effect in many of those cases where it is the salient and correct answer. In most cases, more respondents give the correct answer with "Not necessarily" as the prompt. They suggest also the

possibility that the low scores found by Ennis on invalid items might to some small extent be a product of the test format.

While performance seems to improve on the invalid items taken individually, subjects do not seem to improve quite so much when one looks at the consistency of their performance on groups of items. Ennis characterises mastery of a principle as the correct answering of at least 5 out of the 6 items for that principle; "borderline" performance involves getting 4 of the 6 right. Table 3 shows percentage mastery and borderline performance on the six principles tested for the two groups. As can be seen, there is some improvement in mastery and borderline rates for two of the three principles, but there have been smaller fluctuations between years and there is a trend for performance on *minval* to improve over the years. To some extent this failure of better performance on individual items to translate into mastery is due to the time constraints - Tables 1 and 2 reveal increasing omissions towards the end of the test. But it is also a matter of what is still a hazy grasp of the logic - less than half the subjects get right answers on most of the items.

13th October 1992

REFERENCES

Brandon, E.P. (1990). The Deductive Logical Competence of Non-graduate Caribbean Teachers. ERIC Documentation Service, ED 315 330.

Ennis, R. H. and Paulus, D. H. (1965). *Critical Thinking Readiness in Grades 1-12 (Phase I, Deductive Reasoning in Adolescence)*. Cornell Critical Thinking Project (ERIC Document Reproduction Service No. ED 003 818).

Nolan, C. A. and Brandon, E. P. (1986). Conditional reasoning in Jamaica. Paper given to the Conference on Thinking, Harvard, 1984 (ERIC Document Reproduction Service No. SO 016 755).

Table 1: Item analysis 1990/1 - Valid items

	1990 (N = 537)						1991 (N = 474)						
	Y	N	M	Omit	Right	Diff	Y	N	NN	Omit	Diff	Xsq	*p*
modpon						.83					.84		
q2	426	5	103	3	Y	.79	381	5	83	5	.80	1.24	.74
q4	421	63	44	9	Y	.78	390	42	37	5	.82	3.22	.36
q14	32	447	55	3	N	.83	43	399	29	3	.84	8.49	.04*
q16	510	3	21	3	Y	.95	454	3	17	0	.96	2.76	.43
q19	452	12	67	6	Y	.84	403	7	61	3	.85	1.49	.69
q32	407	24	76	30	Y	.76	357	20	70	27	.75	0.12	.99
modtol						.69					.66		
q5	267	140	108	22	Y	.50	221	110	125	18	.47	5.67	.13
q10	64	389	80	4	N	.72	51	335	87	1	.71	3.68	.30
q15	30	385	113	9	N	.72	32	329	108	5	.69	1.79	.62
q29	419	24	75	19	Y	.78	363	24	69	18	.77	0.36	.95
q33	381	55	67	34	Y	.71	332	33	72	37	.70	5.27	.15
q36	60	363	61	53	N	.68	55	301	77	41	.64	5.49	.14
dissyl						.84					.86		
q1	394	34	102	7	Y	.73	359	23	87	5	.76	1.35	.72
q8	32	463	40	2	N	.86	33	406	35	0	.86	2.17	.54
q13	51	439	41	6	N	.82	46	397	29	2	.84	2.51	.47
q21	496	9	26	6	Y	.92	442	11	16	5	.93	1.86	.60
q25	475	8	42	12	Y	.88	422	15	27	10	.89	4.80	.19
q34	458	7	38	34	Y	.85	401	11	24	38	.85	4.15	.25

Note: the items are grouped by the logical form of the argument tested: modpon is *if p then q, p, so q*; modtol is *if p then q, not q, so not p*; and dissyl is *either p or q, not p, so q*. The difficulty index above each group of items is the mean for that group.

Table 2: Item analysis 1990/1 - Invalid items

	1990 (N = 537)						1991 (N = 474)						
	Y	N	M	Omit	Right	Diff	Y	N	NN	Omit	Diff	Xsq	*p*
ummost						.35					.32		
q3	319	40	172	6	M	.32	309	23	138	4	.29	4.97	.17
q7	234	91	208	4	M	.39	246	72	149	7	.31	9.19	.03*
q12	264	27	242	4	M	.45	262	22	187	3	.39	3.80	.28
q17	248	29	254	6	M	.47	230	36	204	4	.43	3.38	.34
q20	321	29	166	21	M	.31	230	35	195	14	.41	15.46	.00**
q28	393	44	71	29	M	.13	379	30	39	26	.08	8.48	.04*
minval						.40					.50		
q6	203	150	181	3	M	.34	167	78	224	5	.47	27.49	.00***
q11	166	180	183	8	M	.34	139	111	220	4	.46	19.01	.00***
q18	172	87	272	6	M	.51	140	46	284	4	.60	12.70	.01**
q23	178	105	245	9	M	.46	138	71	259	6	.55	8.73	.03*
q26	256	50	214	17	M	.40	190	18	254	12	.54	25.28	.00***
q31	263	52	192	30	M	.36	238	40	169	27	.36	0.51	.92
nonono						.25					.34		
q9	115	271	142	9	M	.26	91	204	168	11	.35	10.74	.01*
q22	180	180	144	33	M	.27	149	129	170	26	.36	10.44	.02*
q24	136	263	110	28	M	.20	93	213	143	25	.30	13.93	.00**
q27	123	280	113	21	M	.21	112	215	126	21	.27	5.85	.12
q30	108	262	137	30	M	.26	93	176	182	23	.38	21.44	.00***
q35	144	180	159	54	M	.30	114	130	171	59	.36	8.32	.04*

Note: see the note to table 1; unmost is the principle *most A are B, all C are A, so some C are B*; minval is *most A are B, most B are C, so most A are C*; nonono is *no A are B, no B are C, so no A are C.*

Table 3: Percentage mastery/borderline performance

1990

	MODPON	MODTOL	DISSYL	UMMOST	NONONO	MINVAL
Mastery	72	47	77	8	7	18
Borderline	15	21	9	15	6	12

1991

	MODPON	MODTOL	DISSYL	UMMOST	NONONO	MINVAL
Mastery	74	43	78	6	9	24
Borderline	13	20	10	12	10	17