

HOW MANY OPTIONS DO MULTIPLE-CHOICE QUESTIONS *REALLY* HAVE?

John R. Dickinson
University of Windsor
MExperiences@bell.net

ABSTRACT

One of the major difficulties—perhaps the major difficulty—in composing multiple-choice questions is the writing of distractors, i.e., the incorrect answer options. Yet distractors play an obviously essential role in determining the effectiveness of a question. Item analysis, and more specifically the analysis of item distractors, is an established tradition. However, only minimal item analysis of ubiquitous multiple-choice question banks accompanying marketing texts has been published and no analyses of distractors. This study describes the extent to which distractors do not, in fact, distract in five such question banks.

INTRODUCTION AND PURPOSE

A defining property of multiple-choice questions is the number of answer options, that number having important implications as described below. The answer options, of course, usually comprise the correct answer plus several distractors (or foils or misleads), i.e., the incorrect answer options. Distractors are the focus of the present study, specifically the number of *effective* distractors (as opposed to the *nominal* number of distractors). The purpose of the study is to examine multiple-choice question banks accompanying several widely adopted (marketing) text books to determine the effective number of distractors.

“The content of an item can be altered radically by changing the distractors, while keeping the correct response the same.” (Cronbach, 1971, p. 454) Accordingly, writing distractors is a challenge for test developers:

- “The major short-comings of multiple-choice questions are, first, the difficulty of writing good distractor options...” (Gregory, 2011, p. 140)
- “When an individual item is being written, the number of potentially meaningful, relevant distractors is far more limited [than the universe of items]; the law of diminishing returns very quickly takes over...the search for good distractors *after* three or four good ones have already been found is likely to be frustrating and fruitless.” (Wesman, 1971, p. 99-98)
- “...preparation of an additional distractor may well require disproportionate additional effort on the part of the item writers.” (Tinkelman, 1971, p. 74)
- “The use of five alternatives is probably the upper limit...due to the difficulty in developing plausible distractors...” (Reynolds & Livingston, 2012, p. 198)

The number of answer options—usually comprising one correct answer plus the distractors—is critical. Fundamentally, the greater the number of options presumably the more difficult it is to answer the question correctly. Distractors that are obviously incorrect effectively reduce the number of response options and, thus, compromise the purpose of testing.

The primary implication of implausible distractors is that measurement of examinees is compromised. “This [measuring students’ levels of comprehension] does not result if the test questions are such that little real knowledge is needed by the testee because of the ease of eliminating ridiculous or remote possibilities in the incorrect choices.” (Weitzman & McNamara, 1945/1946, p. 517)

Operationally, the greater the number of options the less probable that the correct answer can be guessed. “...some test developers suggest using five alternatives [rather than the more common four] because using more alternatives reduces the chance of correctly guessing the answer...” (Reynolds & Livingston, 2012, p. 198). A common practice is to “adjust for guessing” by subtracting points for incorrect answers. Importantly, the adjustment is usually a function of the number of options. For example, a question having five options might score four points for a correct answer, but subtract one point for an incorrect answer. Assuming that guessing equates to choosing one of the five options at random, then the expected value of guessing is zero: $(1/5 \text{ probability of guessing the correct option}) \times (4 \text{ points for a correct answer}) + (4/5 \text{ probability of guessing an incorrect option}) \times (-1 \text{ point for an incorrect answer})$.

However, if one or more of the options is known by the student to be incorrect, he or she will be guessing at random from an effectively smaller number of options, thus resulting in a positive expected value for guessing: “...the guess is not typically made on a totally random basis. It is more reasonable to assume that the testtaker’s guess is based on...the ability to rule out one or more of the distractor alternatives.” (Cohen & Swerdlik, 2010, p. 263) In his item characteristic curve analyses of standard nationally administered tests, Lord (1974, p. 252) found that “...on some items even very low-level examinees may be able to rule out two or three of the possible item responses as incorrect.”

Beyond compromising the measurement of examinees, a second implication of implausible distractors is for item analysis—the refining of multiple-choice questions—i.e., the assessment of questions rather than of examinees. The most basic property of a multiple-choice question is its difficulty. Davis (1951, p. 280) presents a formula for computing item difficulty that is the percent of correct

responses adjusted for chance success, i.e., adjusted for guessing. The adjustment for chance success is based on the number of choices in the item. However, again, where distractors are implausible the effective number of choices is reduced and the Davis-calculated percent correct will be inaccurate.

Distractors that fail to distract, then, do not serve their basic purposes.

“The key [to distractor analysis] is to examine each distractor and ask two questions. First, did the distractor distract some examinees? If no examinees selected the distractor it is not doing its job. An effective distractor must be selected by some examinees. If a distractor is so obviously incorrect that no examinees select it, it is ineffective and needs to be revised or replaced.” (Reynolds & Livingston, 2012, p. 233)

As a general rule, “Make all distractors plausible and attractive to examinees who lack the information or ability tested by the item.” (Wesman, 1971, p. 116). (Later paraphrased by Millman & Greene [1989, p. 353]: “Make all options plausible and attractive to examinees who lack the information or ability referenced by the item.”) “Distractors that are hardly ever chosen are too transparently incorrect and can be omitted or, preferably, replaced.” (Nunnally & Bernstein, 1994, p. 301) and, “...adding distractors that fail to distract cannot improve the utility of the item.” (Wesman, 1971, p. 100) “In multiple-choice tests he [the test writer] learns which distractors (wrong answers) or misleads are not functioning, as shown by their relative unpopularity.” (Guilford, 1954, p. 417)

Multiple-choice question banks are routinely published accompanying virtually all introductory-level business (and other disciplines) textbooks. Yet, despite this ubiquity, very little assessment of those questions has been published. This study examines the distractors of multiple-choice question banks accompanying five texts.

Table 1
Multiple-Choice Question Banks Analyzed

Text	Total Multiple-Choice Questions
Levy, M. & Weitz, B. A. (2012). <i>Retailing management</i> (8 th ed.). [LW 2012]	1194
Solomon, M. R., Zaichkowsky, J. L., & Polegato, R. (2011). <i>Consumer behaviour</i> (5 th Canadian ed.). [SZP 2011]	1152
Levy, M. & Weitz, B. A. (2009). <i>Retailing management</i> (7 th ed.). [LW 2009]	1332
Solomon, M. R., Zaichkowsky, J. L., & Polegato, R. (2008). <i>Consumer behaviour</i> (4 th Canadian ed.). [SZP 2008]	1019
Hawkins, D. I., Mothersbaugh, D. L., & Best, R. J. (2007). <i>Consumer behavior</i> (10 th ed.). [HMB 2007]	1624

Table 2
Sample Questions

Text	Bank Count	Sample Count	Sample as Percent of Bank
LW (2012)	1194	292 ^a	24.5
SZP (2011)	1152	671	58.2
LW (2009)	1332	736	55.3
SZP (2008)	1019	674	66.1
HMB (2007)	1624	958	59.0

^a This relatively small sample of questions is due to the newness of this text edition.

QUESTION BANKS

Data are analyzed for five multiple-choice question banks, three for consumer behavior texts (including two editions of one text) and two for successive editions of a retailing management text. See Table 1. A few questions were deemed invalid in that the correct response was not clear in the text and those questions are excluded. Too, a small number of questions have fewer than the prevalent five options and those questions were excluded.

SAMPLE QUESTIONS

Multiple-choice questions are arranged in the test question banks according to the order in which the question content appears in the textbook. For each examination specific multiple-choice questions were selected on a systematic sampling basis (every 8th or 10th question, with varying starting points) in an attempt to ensure that:

- a cross section of each chapter content was included among the examination questions,
- all three respective exams were of comparable composition, and
- a representative sample of master bank questions was obtained.

The data base of sample questions is summarized in Table 2.

MULTIPLE-CHOICE EXAMS

For all of the courses for which data are available, two midterm exams and one final exam were administered. The exams were not cumulative. The first midterm exam covered about the first third of the chapters (6 or 7 chapters

depending on the specific text), the second midterm covered the middle third of the chapters, and the final exam covered the remaining chapters (5, 6, or 7 chapters). Exams comprised only multiple-choice questions from the relevant master bank. All exams were worth 20 percent of students' final weighted averages for the course. Parameters of exams for each text are presented in Table 3.

ANALYSIS AND RESULTS

The essential purpose of this study is to identify questions having distractors that do not serve their intended purpose. That general purpose is for a distractor to not seem so implausible that examinees who do not know the correct answer can yet dismiss the distractor. This implausibility is reflected in the incidence with which a distractor is selected. Simply put, if a given distractor attracts no or very few responses then the distractor is ineffective.

All questions in this study have four distractors (plus one correct answer option). One or more of those four distractors might be ineffective. In the extreme, for a given question one distractor might attract no responses, two distractors might attract no responses, three distractors might attract no responses, or four distractors might attract no responses (this last being the case where all examinees answer the question correctly). The percents of questions for 0 through 4 distractors selected by no examinees for a given text bank are in the "0%" columns in Table 4.

Less extreme, for a given question one distractor might attract less than or equal to five percent of responses, two distractors might attract less than or equal to five percent of responses, and so on. These results are in the "<=5%" columns of Table 4. And even less extreme results for distractors attracting less than or equal to ten percent of responses are in the "<=10%" columns in Table 4.

Table 3
Exam Parameters

Text	Number of Exams	Mean Questions per Exam	Mean Students per Exam	Mean Score (% Correct)
LW (2012)	6	48.7 (53, 42)	37.2 (39, 35)	67.9 (70.6, 63.4)
SZP (2011)	12	55.9 (60, 47)	41.9 (54, 28)	58.2 (62.8, 54.5)
LW (2009)	12	61.3 (70, 55)	36.2 (49, 27)	67.44 (73.0, 58.5)
SZP (2008)	12	56.2 (60, 48)	39.9 (49, 32)	61.07 (67.1, 57.5)
HMB (2007)	18	53.2 (56, 47)	32.7 (42, 25)	62.74 (69.2, 56.6)

Numbers in parentheses are maximum and minimum, respectively.

DISCUSSION

No published norms for this type of distractor analysis are known to the author. (The present results might be a step toward such norms.) For a great many questions, though, the distractors do not effectively serve their purpose. For each of the five question banks, over half the sample questions had at least one distractor that attracted no responses, the percents of such questions ranging from 53.35 percent to 70.89 percent (Table 5). From 85.01 percent to 91.54 percent had at least one distractor attracting five percent or less of all examinee responses. And from 97.02 percent to 99.16 percent had at least one distractor attracting ten percent or less of responses.

Many basic textbooks in marketing, now multiple editions later, are in the maturity stage of the life-cycle, as are their accompanying question banks. At their inaugural editions, pretesting several hundreds of questions may be infeasible. However, as texts and their question banks are adopted, plentiful item-analysis-relevant data are generated. It should be feasible to revise subsequent editions of question banks accordingly.

LIMITATIONS

Questions analyzed in this study are, of course, samples from their respective published banks of questions and, thus, subject to sampling error. For four of the five question banks, though, over half the population questions were included in the sample (Table 2). The numbers of students, while perhaps typical of non-principles classes, are limited, ranging from averages of 32.7 to 41.9 students per exam/question (Table 3). The universe of course administration and examination parameters seems impossible to formally generalize to. For the present study, though, there is consistency in the administration of the courses and exams (single instructor, common course and exam formats).

The results, too, are pronounced to the point where it seems clear that for the questions analyzed, the effective number of options is *not* the nominal number of options.

REFERENCES

- Cohen, Ronald Jay & Swerdlik, Mark (2010). *Psychological testing and assessment* (7th ed.). Boston: McGraw Hill. ISBN-13: 9780073129099
- Cronbach, Lee J. (1971). Test validation. In Robert L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 443-507. ISBN: 0-8268-1271-6
- Davis, Frederick B. (1951). Item selection techniques. In E.F. Lindquist (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education, 266-328.
- Friedenberg, Lisa (1995). *Psychological testing*. Boston: Allyn & Bacon. ISBN: 0-205-14214-1
- Gregory, Robert J. (2011). *Psychological testing: history, principles, and applications* (6th ed.). Boston: Allyn & Bacon.. ISBN-10: 0-205-78214-0, ISBN-13: 978-0-205-78214-7
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill Book Company.
- Hawkins, Del I., Mothersbaugh, David L., & Best, Roger J. (2007). *Consumer behavior* (10th ed.). Boston: McGraw-Hill Irwin.
- Levy, Michael & Weitz, Barton A. (2012). *Retailing management* (8th ed.). New York: McGraw-Hill Irwin. ISBN-13: 978-0-07-353002, ISBN-10: 0-07-353002-6
- Levy, Michael & Weitz, Barton A. (2009). *Retailing management* (7th ed.). New York: McGraw-Hill Irwin. ISBN-13: 978-0-07-338104-6, ISBN-10: 0-07-338104-7
- Lord, Frederic M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, Vol. 39 (June), 247-264.
- Millman, Jason & Greene, Jennifer (1989). The specification and development of tests of achievement and ability. In Robert L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: American Council on Education and Macmillan Publishing Company, 335-366. ISBN: 0-02-922400-4
- Nunnally, Jum C. & Bernstein, Ira H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill. ISBN: 0-07-047849-X

Table 4
Incidences of Distractor Responses (percent)

Number of Distractors	LW (2012)			SZP (2011)			LW (2009)		
	0%	<= 5%	<= 10%	0%	<= 5%	<= 10%	0%	<= 5%	<=10%
0	29.11 ^a	12.67	1.71	46.65	13.71	2.98	29.76	8.56	0.95
1	34.93 ^b	26.37	9.59	33.38	26.23	14.46	32.20	22.83	10.33
2	21.92	28.42	23.29	15.80	34.13	31.30	23.51	28.26	23.10
3	11.30	21.23	35.62	3.43	18.63	34.72	11.68	27.85	37.09
4	2.74	11.30	29.79	0.75	7.30	16.54	2.85	12.50	28.53
	292 questions			671 questions			736 questions		

a 29.11 percent of 292 questions had 0 distractors attracting 0% of responses

b 34.93 percent of 292 questions had 1 distractor attracting 0% of responses

- Reynolds, Cecil R. & Livingston, Ronald B. (2012). *Mastering modern psychological testing: theory and methods*. Boston: Pearson. ISBN-10: 020548350X, ISBN-13: 9780205483501
- Solomon, Michael R., Zaichkowsky, Judith L., & Polegato, Rosemary (2011). *Consumer behaviour (5th Canadian ed.)*. Toronto: Pearson Prentice Hall. ISBN: 978-0-137-01828-4
- Solomon, Michael R., Zaichkowsky, Judith L., & Polegato, Rosemary (2008). *Consumer Behaviour (4th Canadian ed.)*. Toronto: Pearson Prentice Hall. ISBN-13: 978-0-13-174040-2, ISBN-10: 0-13-174040-7
- Tinkelman, Sherman N. (1971). Planning the objective test. In Robert L Thorndike (Ed.), *Educational measurement (2nd ed.)*. Washington, D.C.: American Council on Education, 46-80. ISBN: 0-8268-1271-6
- Weitzman, Ellis & McNamara, Walter J. (1945/1946). Apt use of inept choice in multiple choice testings. *Journal of educational research*, Vol. 39, 517-522.
- Wesman, A. G. (1971). Writing the test item. In Robert L. Thorndike (Ed.), *Educational measurement (2nd ed.)*. Washington, D.C.: American Council on Education, 81-129. ISBN: 0-8268-1271-6