THE EFFECT OF THE REAL NUMBER OF OPTIONS ON THE DIFFICULTY OF MULTIPLE-CHOICE QUESTIONS

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

ABSTRACT

Despite their ubiquity, published banks of multiple-choice questions have received scant evaluation. The present research investigates the effect of one property of multiple-choice distractors, i.e., incorrect answer options, on question difficulty.

INTRODUCTION

Multiple-choice question banks accompanying many, if not most, introductory-level business textbooks. Despite this, little study has been made of them. Over the twenty-plus years since Dickinson, Faria, & Whiteley (1991) published one such study, only a handful of like researches have ensued. Nonetheless, Dickinson (2012) has evaluated the accuracy with which questions are classified as to level of difficulty and has developed a statistic (2013b), *TaxI*, for measuring this accuracy. The present research investigates the effect of one characteristic of multiple-choice question distractors on the most common item analysis criterion, item difficulty.

ITEM DIFFICULTY

The difficulty of a multiple-choice question is perhaps its most fundamental property. "The first characteristic of item responses is item difficulty." (Haladyna, 2004, p. 207) "One of the most important responsibilities of the test planner is to define the level and the distribution of the difficulties of the items that are to compose the final test." (Tinkelman, 1971, p. 62) Item difficulty is attended by numerous issues including:

- the desired or optimal level of difficulty, whether the exam is to be used for screening purposes (calling for a preponderance of either relatively difficult or relatively easy items) or achievement/discrimination purposes (calling for items of medium and within a limited range of difficulty),
- the incidence of guessing possibly being related to difficulty (i.e., examinees being more likely to guess when an item is difficult),
- sequential testing (in which the difficulty of subsequent items depends on examinee performance),
- and level of discrimination (extremely easy and extremely difficult items having little discriminating ability). (Thorndike, 1971; Tinkelman, 1971)

Presumably the central determinant of the difficulty of a multiple-choice is, or should be, its content, specifically the content of the question stem and of the correct answer option. Distractors, i.e., the incorrect answer options, though, are an integral element of multiple-choice questions. And the characteristics of distractors–specific to the present research their (un) attractiveness– might also affect item difficulty. If this is the case, then the above noted issues attending item difficulty might be at least partially addressed via the incorrect answer options.

DISTRACTORS

Distractors (or foils or misleads) are the incorrect answer options. Perhaps obviously, the purpose of distractors is to provide possible answers for students who do not know the correct answer. This purpose is served, though, only if a given distractor does, in fact, attract some responses.

> "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)

Dickinson (2013a) has shown that for samples of questions from several question banks, this purpose is not served. Across five question banks, the percent of sample questions having at least one distractor attracting no responses ranged from 53.53% to 70.89%. The percent of questions with at least one distractor attracting ten percent or less of total responses ranged from 97.02% to 99.16%.

The effect of distractors that attract few responses might seem to be to make the question easier to answer correctly; students who do not know the answer have fewer options from which to guess. This, however, is not a necessary effect. It is possible that once the item writer has composed one or two effective distractors, the writer does not give the same effort to composing additional distractors. The "effort-intense" distractors, though, may still be sufficient to distract students who do not know the correct answer.

The difficulty of writing distractors is widely recognized:

- "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. 99-100)

- "...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)

In light of the above, investigating empirically the effect of distractors–specifically the inability of some to attract responses –on item difficulty is warranted. That is the purpose of the present study.

DATA

Multiple-choice question banks accompanying five texts were examined. Among the five were two editions of a consumer behavior text plus a second consumer behavior text and two editions of a retailing text. The texts, the total number of multiple-choice questions in the respective banks, and the number of questions sampled from each question bank are reported in Table 1.

Examinations

Providing data for the present analyses were undergraduate courses typically taken in the third year of a student's university program, the courses having as prerequisites two semester-long principles of marketing courses. For each class the first midterm exam covered about the first third of the chapters, the second midterm exam covered about the middle third of the chapters, and the noncumulative final exam covered about the last third of the chapters (Table 2). Each of the exams counted for 20 percent of the students' final course grades.

Exams were scored as the percent of questions answered correctly; no penalty was deducted for incorrect answers. In the very few instances where a question was not answered or multiple answers were given, these were considered to be incorrect answers, both for exam scoring purposes and for the present research. Mean class sizes ranged from 32.7 to 41.9 (Table 2).

Sampling Method

Multiple-choice questions are arranged in the test question bank 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. This systematic sampling approach was an attempt to ensure that:

- a cross section of each chapter content was included among the examination questions,
- all respective midterm and final examinations were of comparable composition, and
- a representative sample of the text bank questions was obtained.

Counts of test bank and sample questions are reported in Table 1. All questions analyzed had five options: the correct answer plus four distractors.

ANALYSIS

The purpose of this research is to determine whether distractors that attract few or no responses affect item difficulty. Item difficulty was measured as the percent of students answering the question correctly. Percent correct is a near-universally prescribed measure of item difficulty (Anastasi & Urbina, 1997, p. 173; Gregory, 2011, p. 141; Guilford, 1954, p. 418; Gulliksen, 1950, p. 366; Henrysson, 1971, p. 139; Nunnally & Bernstein, 1994, p. 301).

Distractors attracting no or few responses were measured in three ways:

- The number of distractors attracting zero responses.
- The number of distractors attracting less than or equal to 5

Table 1
Bank and Sample Question Counts

Text	Total Questions	Sample Questions (percent of total)
Levy & Weitz (2012, LW), *Retailing Management*, Eighth Edition	1211	624 (51.5)
Solomon, Zaichkowsky, & Polegato (2011, SZP), *Consumer Behaviour*, Fifth Canadian Edition	1148	671 (58.4)
Levy & Weitz (2009, LW), *Retailing Management*, Seventh Edition	1332	736 (55.3)
Solomon, Zaichkowsky, & Polegato (2008, SZP), *Consumer Behaviour*, Fourth Canadian Edition	1019	674 (66.1)
Hawkins, Mothersbaugh, & Best (2007, HMB), *Consumer Behavior*, Tenth Edition	1624	958 (59.0)

percent of total responses (the total including correct responses).

- The number of distractors attracting less than or equal to 10 percent of total responses.

Bivariate regression analysis was used to estimate the effect of distractors having sparse responses on item difficulty. It was anticipated that the regression slope would be positive; the greater the number of distractors having sparse responses, the higher the percent correct. (Percent correct is actually an inverse measure of difficulty.)

Regressions were carried out for each question bank separately and for each of the three measures of sparse responses itemized above. The standardized slope coefficient (β), of course, is equal to the Pearson correlation between the two variables.

RESULTS

Standardized slope coefficients of the simple regressions of percent correct on numbers of distractors attracting no or few responses are presented in Table 3. As anticipated, all of the slopes are positive indicating that the greater the presence of sparse distractors the higher the percent of students answering the item correctly. All of the slopes (equal to the Pearson correlation) are statistically significant (one-tail p<.001).

Also as would be expected, as the measure of "sparse" becomes broader (from 0% to ≤5% to ≤10%), i.e., the number of distractors qualifying as sparse increases, the slopes (correlations) increase materially.

Perhaps the most dramatic results are the R^2 values presented in Table 3. The number of distractors attracting zero responses explains between 29.25 (SZP 2011) and 36.65 (LW 2012) percent of the variation in item difficulty. The number of

Table 2
Exams and Students

	Text Chapters					
Text	First Exam	Second Exam	Third Exam	Questions per Exam *	Students per Exam *	Score *
LW (2012)	1-6	7-12	13-18	52.0	38.0	69.5
SZP (2011)	1-6	7-12	13-17	55.9	41.9	58.2
LW (2009)	1-6	7-12	13-19	61.3	36.2	67.4
SZP (2008)	1-6	7-12	13-17	56.2	39.9	61.1
HMB (2007)	1-7	8-14	15-20	53.2	32.7	62.7

* mean

Table 3
Standardized Regression Slopes *
(R^2)

Text	0% of Responses	≤ 5% of Responses	≤ 10% of Responses
LW (2012)	0.6054 (0.3665)	0.7340 (0.5387)	0.8268 (0.6836)
SZP (2011)	0.5408 (0.2925)	0.7122 (0.5073)	0.7495 (0.5617)
LW (2009)	0.5951 (0.3541)	0.7245 (0.5249)	0.8070 (0.6512)
SZP (2008)	0.5522 (0.3049)	0.6932 (0.4805)	0.7864 (0.6184)
HMB (2007)	0.5470 (0.2993)	0.6881 (0.4735)	0.7817 (0.6111)

* All one-tail p-values < .001

distractors attracting 10 percent or fewer responses explains between 56.17 (HMB 2007) and 68.36 (LW 2012) percent of the variation in item difficulty.

Addressing the purpose of this study, the presence of distractors that do not, in fact, distract has a very material effect on item difficulty. The consistency of results across the several question banks reinforces this conclusion.

DISCUSSION

The main implication of this study is that distractors–specifically distractors that attract no or few responses–can materially affect item difficulty. In turn, respecting the issues related to item difficulty itemized above, item writers might attend to distractors as well as to the question stem and correct answer option.

The results of this research, of course, do not necessarily hold for all published banks of multiple-choice questions. There exist any number of guides for writing multiple-choice questions (Gregory, 2011, p. 140; Haladyna, 2004; Reynolds & Livingston, 2012, pp. 197-202; Wesman, 1971). The many different item writers, though, are not necessarily in lock-step with those guides. Nor do those guides encompass relevant human characteristics of item writers such as subject expertise, ingenuity, empathy with target students, straightforward expression, and so on.

The consistency of the results across the several test banks (those of two editions no doubt having several duplicated questions), though, suggests some reliability of the findings.

Data for replicating this research are plentiful and easily obtained. Such replication might further support (or not) the essential result of this study. Too, publishers might carry out similar investigations of their question banks. Many texts publish periodic editions (LW being in its eighth edition, SZP being in its fifth edition, and HMB now being in its thirteenth edition). Refining the distractors (or other properties) of multiple-choice questions from edition to edition would soon see improved question banks, of benefit to publishers specifically and pedagogy generally.

REFERENCES

Anastasi, Anne & Urbina, Susana (1997). *Psychological testing*, Seventh Edition. Upper Saddle River, NJ: Prentice Hall.

Dickinson, John R. (2013a). How many options do multiple-choice questions *really* have? In Marian Boscia (Ed.), *Developments in Business Simulation and Experiential Learning*, Association for Business Simulation and Experiential Learning, Vol. 40, 171-175. (Reprinted from *Bernie Keys Library*.)

Dickinson, John R. (2013b). *TaxI*: a statistic describing the accuracy of multiple-choice question difficulty classifications. In Krzysztof Kubacki (Ed.), *2013 Proceedings of the Annual Conference of The Academy of Marketing Science*.

Dickinson, John R. (2012). A taxonomy assessment and item analysis of a consumer behaviour Multiple-choice question bank. In Lorne Sulsky (Ed.), *Administrative Sciences Association of Canada Conference Proceedings*, Vol. 33, 6-31.

Dickinson, John R., Faria, A. J., & Whiteley, T. Richard (1991). An empirical investigation of the validity of McCarthy's multiple-choice question classification matrix. *Journal of Marketing Education*, Vol. 13 (Summer), 54-66.

Gregory, Robert J. (2011). *Psychological testing: history, principles, and applications*, Sixth Edition. Pearson. ISBN -10: 0205782140, ISBN-13: 9780205782147

Guilford, J. P. (1954). *Psychometric methods*, Second Edition. New York: McGraw-Hill Book Company.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Haladyna, Thomas M. (2004). *Developing and validating multiple-choice test items*, Third Edition. New York: Routledge. ISBN: 0-8058-4661-1

Hawkins, Delbert I., Mothersbaugh, David L., & Best, Roger J. (2007). *Consumer behavior*, Tenth Edition. Boston: McGraw-Hill Irwin. ISBN-13: 978-0-07-310137-8, ISBN-10: 0-07-310137-0

Henrysson, Sten (1971). Gathering, analyzing, and using data on test items. In Robert L. Thorndike (Ed.), *Educational Measurement*, Second Edition. Washington, D.C.: American Council on Education, 130-159.

Levy, Michael & Weitz, Barton A. (2012). *Retailing management*, Eighth Edition. 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*, Seventh Edition. New York: McGraw-Hill Irwin. ISBN-13: 978-0-07-338104-6, ISBN-10: 0-07-338104-7

Nunnally, Jum C. & Bernstein, Ira H. (1994). *Psychometric theory*, Third Edition. New York: McGraw-Hill, Inc..

Reynolds, Cecil R. & Livingston, Ronald B. (2012). *Mastering modern psychological testing: theory and methods*. Pearson. ISBN-10: 020548350X, ISBN-13: 9780205483501

Solomon, Michael R., Zaichkowsky, Judith L., & Polegato, Rosemary (2011). *Consumer behaviour*, Fifth Canadian Edition. Toronto: Pearson Prentice Hall. ISBN: 978-0-137-01828-4

Solomon, Michael R., Zaichkowsky, Judith L., & Polegato, Rosemary (2008). *Consumer behaviour*, Fourth Canadian Edition. Toronto: Pearson Prentice Hall. ISBN-13: 978-0-13-174040-2, ISBN-10: 0-13-174040-7

Thorndike, Robert L. (1971). Educational measurement for the seventies. In Robert L. Thorndike (Ed.), *Educational Measurement*, Second Edition. Washington, D.C.: American Council on Education, 3-14. ISBN: 0-8268-1271-6

Tinkelman, Sherman N. (1971). Planning the objective test. In Robert L. Thorndike (Ed.), *Educational Measurement*, Second Edition. Washington, D.C.: American Council on Education, 46-80. ISBN: 0-8268-1271-6

Wesman, A. G. (1971). Writing the test item. In Robert L. Thorndike (Ed.), *Educational Measurement*, Second Edition. Washington, D.C.: American Council on Education, 81-129. ISBN: 0-8268-1271-6