The Unsuitability of Goodman and Kruskal's Lambda Measure of Association for *TaxI* Analysis of Multiple-Choice Question Difficulty Taxonomies

John R. Dickinson
University of Windsor, Canada
MExpeeriences@bell.net

ABSTRACT

TaxI analysis of published multiple-choice question bank difficulty taxonomies produces classification matrices relating measured or observed question difficulty to published difficulty level, i.e., the accuracy of thee published taxonomy. Where there is a preponderance of questions in one of the published categories, an anomaly in the Goodman and Kruskal lambda measure of association renders it unsuitable for TaxI classification matrices The present study explains that anomaly and illustrates its unsuitability are explained.

INTRODUCTION

Banks of multiple choice questions are ubiquitous, accompanying virtually every introductory marketing and other business textbook. Questions ae usually classified into three levels of difficulty–Easy, Medium, Hard–as well as on other dimensions (e.g., skill, Bloom, AACSB). Recently, Dickinson (2013) has introduced the Taxonomy Index (*TaxI*) that describes the accuracy of the published difficulty classifications. One product of a *TaxI* analysis is a three by three classification matrix the rows being published classified difficulty with the columns being observed or measured difficulty. (Table 1) For cross-tabulations, Reynold (1984, 30-71) presents several measures of association : "A measure of association is a numerical index summarizing the strength or degree of relationship in a two-dimensional cross-classification." (p. 20) Among the measures is Goodman and Kruskal's (1954, G-K) lambda which, "...rests on very straightforward definitions of prediction error." (Reynolds, 1984, p. 51). That is, given the value of one variable to what degree is the error in predicting the second variable decreased compared with no knowledge of the first variable. Lambda or λ, is referred to as the "index of predictive association" (Iacobucci and Churchill, 2010, p. 381). "...lambda measures the percentage improvement in predicting the value of the dependent variable, given the value of the independent variable. Lambda...varies between 0 and 1. A value of 0 means no improvement in prediction. A value of 1 indicates that the prediction can be made without error." (Malhotra, 2010, pp. 469-470)

TABLE 1
CLASSIFICATION MATRIX FOR *TAXI* ANALYSIS OF 612 QUESTIONS SYSTEMATICALLY RANDOMLY SAMPLED FROM 1210 QUESTIONS IN THE PUBLISHED BANK ACCOMPANYING LWG (2014)

		OBSERVED (row percent / count)			
		Easy	Medium	Hard	Total
CLASSIFIED	Easy	44.04 96	56.50 121	0.45 1	218
	Medium	31.12 117	64.34 242	4.52 17	376
	Hard	27.78 5	72.22 13	0.00 0	18
	Total	218	376	18	512
TaxI = 0.552				C_{pro} = 0.505	
TaxI is the proportion of questions correctly classified. C pro is the proportion of questions that would be classified correctly by chance, i.e., without knowledge of the published taxonomy.					
Levy, Weitz, and Grewal (2014), *Retailing Management* Ninth Edition					

In the context of the present research this is predicting observed or measured question difficulty level from the published difficulty level.

TaxI measures the accuracy of the published taxonomy. The strength of association between published and observed difficulty is a complementary property. An anomaly in Goodman and Kruskal's lambda (explained below), though, is evident in most of the question banks (Table 4) analysed for this research (Table 2), making it unsuitable for *TaxI* analyses.

TABLE 2
TEXTBOOKS, PUBLISHED QUESTIONS, AND SAMPLE QUESTIONS

Text	Total Questions	Sample Questions
Levy, Weitz, and Grewal (2019), *Retailing Management*, Tenth Edition	1157 [a]	479 [a]
Cateora, Gilly, Graham, and Money (2016, Cateora), *International Marketing*, Seventeenth Edition	1178	425
Levy, Weitz, and Grewal (2014), *Retailing Management* Ninth Edition	1210	512
Levy and Weitz (2012), *Retailing Management*, Eighth Edition,	1190	624
Solomon, Zaichkowsky, and Polegato (2011, SZP), *Consumer Behaviour*, Fifth Canadian Edition	1148	671
Levy and Weitz (2009, LW), *Retailing Management*, Seventh Edition	1332	736
Solomon, Zaichkowsky, and Polegato (2008), *Consumer Behaviour*, Fourth Canadian Edition	1019	674
Hawkins, Mothersbaugh, and Best (2007, HMB), *Consumer Behavior*, Tenth Edition	1624	958

a 1157 questions are in the published bank. Of these, a systematic random sample of 479 questions was drawn.

HOW GOODMAN AND KRUSKAL'S LAMBDA WORKS AND AN ANOMALY

The G-K lambda is of the proportional reduction in error (PRE) measure genre. Simply put, the best prediction of the specific category of the dependent variable (here the observed or measured difficulty) without knowledge of the independent (row) variable is the modal category of the dependent variable. The total prediction error is the sum of the counts in the nonmodal categories, error(w/o row).

Knowing the category of the independent (i.e., row or published difficulty) variable, the best perdition is the column in which the highest cell in the row resides. The "error" is the sum of the counts in the remaining cells in the row. Gathered across all the rows is the total prediction error given the independent variable, i.e., the row. error|row. G-K's lambda equals [error(w/o row) - error|row] / error(w/o row).

It is possible, though, for the chi-square test of the null hypophysis of independence (of the row and column variables) to be highly significant with a small p-value supporting the conclusion that the two variables *are* associated. and the G-K lambda to equal zero.

This anomaly in G-K lambda occurs when the number of observations in one column is much greater than the numbers in the other columns. This can lead to the greatest cell count in each row being in the same column. That is, the predicted column is the same for each row. This marked predominance of questions in one difficulty level over the other two difficulty levels is evident in the banks analysed for the present research. As reported in Table 3, for all of the banks the percentage of questions in one of the classified difficulty categories is materially greater than in the other two categories. For all but one of the banks, that predominant category is Medium.

As a consequence of one published difficulty level containing a predominance of questions, for five of the eight question

TABLE 3: PREDOMINANT CLASSIFIED DIFFICULTY LEVEL
Bank percentage (count) of questions in each classified difficulty taxonomy level
(bold indicates difficulty category with the greatest number of questions)

Text	Easy	Medium	Hard
LWG (2019), Tenth	36.65% (424)	**60.24% (697)**	3.11% (36)
Cateora (2016), Seventeenth	**57.98% (683)**	36.84% (434)	5.18% (61)
LWG (2014), Ninth	34.46% (417)	**62.98% (762)**	2.56% (31)
LW (2012), Eighth	30.76% (366)	**58.32% (694)**	10.92% (130)
SZP (2011), Fifth	25.44% (292)	**49.74% (571)**	24.83% (285)
LW (2009), Seventh	31.61% (421)	**56.23% (749)**	12.16% (162)
SZP (2008), Fourth	25.02% (255)	**50.05% (510)**	24.93% (254)
HMB (2007), Tenth	14.72% (239)	**70.69% (1148)**	14.59% (237)

banks lambda equals zero. This, despite the null hypothesis of independence being decidedly rejected (χ^2 p-value □0.001 for seven of the eight question banks, Table 4).

TABLE 4:
GOODMAN AND KRUSKAL'S LAMBDA, *TAXI* AND CRAMER'S V

Text	Sample Questions	χ^2 p-Value	G-K Lambda	*TaxI*	Crsmer's V
LWG (2019) 10th	41.002% = 479/1157	0.000	0.0258	58.246	0.135
Cateora (2016), 17th	36.0781% = 425/1178	0.0672	**0.0000**	49.647	0.102
LWG (2014), 9th	50.5785% = 612/1210	0.001	**0.0000**	55.229	0.119
LW (2012), 8th	52.4370% = 624/1190	0.001	**0.0000**	49.199	0.120
SZP (2011), 5th	58.4495% = 671/1148	0.000	0.0673	50.969	0.263
LW (2009), 7th	55.2553% = 736/1332	0.000	**0.0000**	51.495	0.172
SZP (2008), 4th	66.1433% = 674/1019	0.000	0.0360	49.258	0.264
HMB (2007), 10th	58.9902% = 958/1624	0.000	**0.0000**	61.169	0.213

CONCLUSION

In putting forth their lambda measure Goodman and Kruskal (1954, p. 740, italics theirs) observe, "The fact that an excellent test of independence may be based on χ^2 does not at all mean that χ^2, or some simple function of it, is an appropriate *measure* of degree of association Among the several measures of association presented by Reynolds (9184, pp. 30-71) is Cramer's V:

$$\text{Cramer's V} = \frac{\chi^2}{\sqrt{\text{sample size X minimum rows-1)(columns-1)}}}$$

Cramer's V varies between 0 and 1. Values for the eight classification matrices are presented in Table 4. In light of the highly significant relationship in all of the *TaxI* classification matrices and the proportions of correctly classified questions being greater than chance as measured by the *TaxI* statistic, the materially greater than zero values of Cramer's V are a more informative complement to *TaxI* analysis than the anomalous zero values of Goodman and Kruskal's lambda.

REFERENCES

Cateora, Philip R., Gilly, Mary C., Graham, John L., and Money, R. Bruce (2016), *International Marketing*, Seventeenth Edition (New York: McGraw-Hill Education). ISBN: 978-0-07-7842166-1, MHID: 0-07-784216-2

Dickinson (2013),"*TaxI*: A Statistic Describing the Accuracy of Multiple-Choice Question Difficulty Classifications," in Kubacki, Krzystof (editor), *Proceedings of the 2013 Academy of Marketing Science Annual Conference*, Vol. XXXVI, p. 337.

Goodman, Leo A. and Kruskal, William H. (1954), "Measures of Association for Cross Classifications," *Journal of the American Statistical Association* Vol. 49, No. 268 (December), 732-764.

Hawkins, Del I., Mothersbaugh, David L., & Best, Roger J. (2007). *Consumer Behavior*, Tenth Edition. Boston: McGraw-Hill Irwin.

Iacobucci, Dawn and Churchill, Gilbert A., Jr. (2010), *Marketing Research: Methodological Foundations*, Tenth Edition (Mason, OH: South-Western Cengage Learning). Student Edition ISBN 13: 978-0-324-35995-4; Student Edition ISBN 10: 0-324-35995-0

Levy, Michael, Weitz, Barton A., & Grewal, Dhrus (2019), *Retailing Management*, Tenth Edition (New York: McGraw-Hill Education). ISBN: 978-1-260-08476-4, MHID 1-260-084-76-0

Levy, Michael, Weitz, Barton A., and Grewal, Dhruv (2014), *Retailing Management*, Ninth Edition (New York: McGraw-Hill Education). ISBN 978-0-07-802899-1, MHID 0-07-802899-1

Levy, Michael and 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 (2009). *Retailing Management*, Seventh Edition. New York: McGraw-Hill Irwin.

Malhotra, Naresh K. (2010), *Marketing Research: An Applied Orientation*, Sixth Edition (Boston: Prentice Hall). ISBN: 978-0-13-608543-0

Reynolds, Henry T. (1984), Analysis of Nominal Data (Beverly Hills, CA: Sage Placations, Inc.). ISBN: 0-8039-0653-6

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