Journal of Research Methodology, Volume 16, Number 4 (October - December 2003)

วารสารวิธีวิทยาการวิจัย ปีที่ 16 ฉบับที่ 4 (ตุลาคม - ธันวาคม 2546)

Choice-Agreement Index and Its Application to Item Analysis

Teara Archwamety

ABSTRACT

The present paper proposes an index used to measure the degree of agreement among a group of N choosers in choosing from K alternatives—the Choice-Agreement Index. The index is based on the notion of agreement pairs and disagreement pairs. It has a range of values from zero to one. The derivation of the basic formula for the index and its simplified version suitable for computation are described. An application of the index in analyzing the distractibility of the distractors in a test item is also suggested. This gives rise to the concept of collective distractibility in test item analysis. Twenty-seven items from a recent final exam were selected for a demonstration of this application. The newly suggested collective distractibility index and the classical item difficulty index were computed for each of the 27 test items. The correlation between the two indexes was found to be 0.49, $p < .01$.

Measuring agreement between and among people has long been studied under the popular general topics of correlation and reliability. Measures of interpersonal agreement have been well established for *interval scale* in concepts such as interrater reliability (see Fan & Chen, 2000; Epstein, Cullinan, Harniss & Ryser, 1999), scorer or interjudge reliability (see Anastasi, 1982, p. 117; TenBrink, 1974, p. 39), intraclass correlation (see Robinson, 1957; Winer, 1971, p. 286), and coefficient of agreement A (Robinson, 1957). While the Pearson correlation (Galton, 1888) could be used to compute interpersonal agreement between two judges, or two raters, or two scorers, Hoyt's reliability by analysis of variance (Hoyt, 1941) could be used in the cases of more than two raters or scorers. Measures of interpersonal agreement have also been well established for *ordinal scale* such as the Spearman rank correlation (Spearman, 1904), and the Kendall coefficient of concordance W (Kendall, 1948). However, the same could not be said for *nominal scale* or variable. Light (1971) commented on the lack of investigation in this area. Although a few measurement techniques of measuring interpersonal agreement at the nominal level exist (see Light, 1971, 1973), they are about the interpersonal agreement on how to assign a set of objects to different categories. For example, how much do two people (two-dimensional) or three people (three-dimensional) agree on how to assign a set of objects to different categories? In this paper, we are more interested in a different type of interpersonal agreement. We are interested in how a group of people (usually more than two or three people) assign themselves to different choices or categories. How much do they agree on those choices? Despite the lack of investigation in this type of interpersonal agreement, this type of agreement widely exists in real life. Choices are often presented to a group of people. In shopping, consumers have to choose among different brands. In education, students have to choose among choices in a multiple-choice test. In politics, voters have to choose among different candidates. An index is very much in need to measure *choice agreement* among choosers. The following is a more concrete example of the issue the author would like to deal with in the present paper.

As a result of a survey conducted at a college, 194 returned questionnaires showed that 22 personnel favored disability income insurance with TIAA, 159 with NSEA and 13 were undecided. How much agreement did the personnel have on the choices? One way to get a feel of the answer to this question is to display the distribution of the choices made and compute the percentages, as shown in Table 1.

However, it would be convenient to have a single number (index) indicating the degree to which the personnel agreed on the three choices made in the same way it is convenient to represent the variability of a set of numbers by SD, the Standard Deviation.

The purpose of this paper is to present the development of an index used to measure how much agreement a group of people have in choosing from a finite number of alternatives--the *Choice-Agreement Index.* In addition, an example of the index's application in the area of test item analysis will be demonstrated.

Methods and Procedures

Development of the Index

Let's start with a simple hypothetical example. Nine people are to choose or vote yes or no or undecided on a particular issue. The result looks like this:

	Frequency
Yes	5
No	3
Undecided	1

How much is the agreement?

We could start out by considering how many possible pairs of choosers (P_t) there are. Then we consider how many agreement pairs (P_a) there are, and how many disagreement pairs (P_d) there are. An *agreement pair* is one in which members of the pair choose the same alternative. A *disagreement pair* is one in which members of the pair choose different alternatives. An index indicating how much agreement there is might be:

$$Index_1 = \frac{\text{number of agreement pairs}}{\text{number of all possible pairs}} \quad \text{or}$$

$$Index_1 = \frac{P_a}{P_t} \qquad (1)$$

Of course, the number of all possible pairs minus the number of agreement pairs is the number of disagreement pairs, or

$$P_d = P_t - P_a \tag{2}$$

In our hypothetical example, with 9 choosers, we have

$$P_t = \binom{9}{2} = \frac{9!}{2!7!} = 36 \quad \text{possible pairs}$$

Since 5 choosers choose Yes, we have

$$\binom{5}{2} = \frac{5!}{2!3!} = 10 \quad \text{agreement pairs on Yes.}$$

Since 3 choosers choose No, we have

$$\binom{3}{2} = \frac{3!}{2!1!} = 3 \quad \text{agreement pairs on No.}$$

Since only 1 chooser chooses *Undecided,* we have 0 agreement pair on *Undecided.*

Thus the total number of agreement pairs

$$P_a = 10 + 3 + 0 = 13.$$

Therefore, the index indicating the choice-agreement in our hypothetical example is:

$$Index_1 = \frac{P_a}{P_t} = \frac{13}{36} = 0.36$$

A flaw in the above way of calculating the index is that the lower limit of the index does not reach 0 although the above limit reaches 1. If all choosers agree to say *No* (*or Yes or Undecided*), we have $P_a = P_{a.max} = P_t$ and the index is, therefore, 1.00. However, with 3 alternatives and 9 choosers as in our hypothetical example, we can never have P_a = 0 and, therefore, cannot have a value of the index = 0.00. P_a can never be 0 in this case because *however divided* the decision is, there will still be more than one chooser in *at least one* same alternative. What is the lowest possible number of agreement pairs ($P_{a.max}$) we can get in our hypothetical example? The answer is $3 \times \binom{3}{2} = 9$ pairs.

It is also possible that this $P_{a.max}$ will vary from case to case, depending on how many choosers (N) and how many alternatives (K) there are, and therefore will make the comparison of indexes among cases difficult.

One way to get around this problem is to normalize the range $P_{a.min}$ to $P_{a.max}$. That is to say we subtract $P_{a.min}$ from P_a and divide the result by the difference between $P_{a.max}$ and $P_{a.min}$. Doing this, we arrive at an index:

$$Index_2 = \frac{P_a - P_{a.min}}{P_{a.max} - P_{a.min}} \quad \text{or since } P_{a.max} = P_t$$

$$Index_2 = \frac{P_a - P_{a.min}}{P_t - P_{a.min}} \tag{3}$$

This index will range from 0 to 1 regardless of how many choosers (N) and how many decision alternatives (K) there are.

To simplify Equation 3, consider the fact that whenever the number of agreement pairs is at the lowest possible, the number of disagreement pairs is at the highest possible. Therefore, from Equation 2, we have

$$P_{d.max} = P_t - P_{a.min} \quad \text{or}$$

$$P_{a.min} = P_t - P_{d.mix} \tag{4}$$

Also from Equation 2, we have

$$P_a = P_t - P_d \tag{5}$$

Substituting Equations 4 and 5 into Equation 3, we get

$$Index_2 = \frac{P_t - P_d - (P_t - P_{d.max})}{P_t - (P_t - P_{d.max})} = 1 - \frac{P_d}{P_{d.max}} \tag{6}$$

It can be shown mathematically through the use of differential calculus (see Appendix for complete proof) that

$$P_{d.max} = \frac{(K-1)N^2}{2K}$$

Thus, in the case of 3 alternatives or K = 3,

$$P_{d.max} = \frac{N^2}{3} \tag{7}$$

In our hypothetical example, P_a was previously computed to be 13 and P_t to be 36. Therefore, P_a = 36 - 13 = 23 (see Equation 2). From Equation 7, we have $P_{a.max}$ = 9 x 9 / 3 = 27. Therefore, for our hypothetical example,

$$Index_2 = 1 - \frac{23}{27} = 0.15.$$

Generalization to N Choosers and K alternatives

For K alternatives the distribution of choosers through the alternatives is n_1, n_2, n_3, ... n_K and the number of disagreement pairs(P_d) is $n_1n_2+n_1n_3+$ ... $+n_1n_K$ + $n_2n_3+n_2n_4+$ ... $+n_2n_K$ + ... $+n_{K-1}n_K$ or, in short, $\sum\sum_{i<j} n_i n_j$ (i = 1, K−1; j = 2, K).

For any K alternatives and N choosers, it can be shown, through the use of differential calculus (see Appendix for complete proof), that $P_{d.max}$ occurs when N is distributed *evenly* through the alternatives or, in other words, each n = N/K. This corresponds to intuitive thinking! The decision is most split when choosers fill in all the alternatives equally. This means the total number of disagreement pairs in this case is

$$P_{d.max} = \binom{K}{2}\left(\frac{N}{K}\right)^2 = \frac{(K-1)N^2}{2K} \tag{8}$$

Therefore, the computation formula for Index_2 in a general case is

$$Index_2 = 1 - \frac{\sum\sum_{i<j} n_i n_j}{\binom{K}{2}\left(\frac{N}{K}\right)^2} = 1 - \frac{2K\sum\sum_{i<j} n_i n_j}{(K-1)N^2} \tag{9}$$

Comments on $Index_2$ (Equation 9)

Index_2 as shown in Equation 9 has the disadvantage of being too sensitive as the index approaches 1. For example, in the case of 2 alternatives and 6 choosers, there are 4 possible different levels of decision patterns as shown in Table 2.

Index_2's computed for these 4 levels are 1.00, 0.44, 0.11, and 0.00 respectively. Evidently, the distance between levels 1 & 2 is greater than between level 2 & 3 and that between 2 & 3 is greater than between levels 3 & 4. This pattern of unequal distances between adjacent levels is similar for cases with more alternatives and more choosers. To correct for this deficiency without upsetting the order of the different levels, we could perform a square root on $Index_2$. When this is done, we have the values 1.00, 0.66, 0.33, and 0.00 for the 4 different levels respectively.

The author, therefore, proposes that a proper index for measuring *choice-agreement* be:

$$I_{CA} = \sqrt{1 - \frac{2K \sum\sum_{i<j} n_i n_j}{(K-1)N^2}} \quad (10)$$

where N = the number of choosers

K = the number of alternatives

and i goes from 1 to K−1

j goes from 2 to K

Using this formula, the Choice-Agreement Index (I_{CA}) for the decision data cited in the introduction section of this paper (see Table 1) is:

$$\sqrt{1 - \frac{2(3)[(22)(159) + (22)(13) + (159)(13)]}{(3-1)\ 194^2}}$$

$= 0.73.$

Simplifying the Computation of the Choice-Agreement Index (I_{CA})

Although the computation formula of the Choice-Agreement Index (I_{CA}) as given in Equation 10 does not look too complicated, it does become cumbersome when the number of alternatives (K) increases. The summation of all possible pairs of alternatives becomes lengthy. Fortunately, it turns out that Equation 10 can be further simplified.

To simplify Equation 10, start by noting that:

$$(n_1 + n_2 + \ldots + n_K)^2 = n_1^2 + n_2^2 + \ldots + n_K^2 + 2 \sum\sum_{i<j} n_i n_j = N^2$$

Therefore,

$$\sum_{i=1}^{K} \frac{n_i^2}{N^2} = 1 - \frac{2 \sum\sum_{i<j} n_i n_j}{N^2} \quad (11)$$

Now from Equation 10:

$$I_{CA} = \sqrt{1 - \frac{2K \sum\sum_{i<j} n_i n_j}{(K-1)N^2}}$$

$$= \sqrt{1 - \frac{\left(2K \sum\sum_{i<j} n_i n_j\right) / N^2}{1-1/K}}$$

(by dividing K and N^2 into both numerator and denominator)

$$= \sqrt{\frac{-1/K + 1 - \left(2K \sum\sum_{i<j} n_i n_j\right) / N^2}{1-1/K}}$$

$$= \sqrt{\frac{\sum_{i=1}^{K} \frac{n_i^2}{N^2} - 1/K}{1-1/K}}$$

(by Equation 11)

$$= \sqrt{\frac{\left(\sum_{i=1}^{K} \frac{n_i^2}{N^2}\right) / K - (1/K)^2}{(K-1)/K^2}}$$

(by dividing K into both numerator and denominator)

At this point, note that the numerator is actually the population *variance* of n_1/N, n_2/N, n_3/N, ... n_K/N, with the arithmetic mean of 1/K. Therefore,

$$I_{CA} = \sqrt{K \frac{K}{K-1} \sigma^2}$$

Since K/(K-1) is the conversion factor of a population variance (σ^2) into a sample variance (s^2), we finally have:

$$I_{CA} = \sqrt{Ks^2} \qquad (12)$$

where s^2 is the *sample* variance of n_1/N, n_2/N, n_3/N, ... n_K/N. Because the *sample* *variance* function key is available in many popular low cost calculators, the task of calculating the Choice-Agreement Index (I_{CA})is greatly simplified.

Significance Testing

If two independent random samples of choosers are drawn from their respective populations presented with K choices, how do we determine the significant difference between the two choice-agreement indexes computed for the two samples? One possible solution is to note that the ratio of the two indexes from Equation 12 above yields $\sqrt{Ks_a^2} / \sqrt{Ks_b^2} = s_a/s_b$ the square of which (s_a^2/s_b^2) is well-known to have an F distribution with K-1 and K-1 degrees of freedom. This method of testing significant difference between two choice-agreement indexes would require a large K since the population distribution of choices is unlikely to be normal (see Rosner, 2000, p. 293; Hays, 1973, p. 451). The merit of this method, however, is that it is independent of the number of choosers. Note that the value of the choice-agreement index in Equation 12 depends on the number of choices or alternatives (K), NOT on the number of choosers (N).

In the case of small K, testing significant difference between two choice-agreement indexes could be performed via the standard chi-square test. The distribution of choosers over the K choices in the first independent sample could be compared with that in the second sample. This method of testing significant difference, however, would require a relatively large number of choosers. Moreover, it is also subjected to the popular rule of minimum expected frequency—10 in the case of two choices (single degree of freedom) or 5 in the case of three or more choices (see Hays, 1973, p. 736).

An Application in Item Analysis

A possible application of the Choice-Agreement Index is in the area of test item analysis. Classical methods of test item analysis involve (a) item difficulty index, (b) item discrimination index, and (c) item distractor index. While the item difficulty index and item discrimination index are popular, item distractor index is relatively less used (French, 2001). The distractibility of a wrong choice (also called a decoy or distractor) could simply be measured by the percentage of proportion of students choosing it. Brown (1976, p. 280), for example, suggested that a distractor should be eliminated unless two percent of the students select it. Such a percentage could be referred to as an individual distractor index or, more appropriately, an *individual distraction* index. In a typical multiple-choice test item with three distractors, we may have three individual distraction indexes such as 0.10, 0.15, and 0.20. If we sum these three individual distraction indexes, we would have something like a *total item distraction* index. In the example given, this total item distraction index would be 0.10+0.15+0.20=0.45.

The total *item distraction* index mentioned in the previous paragraph is not a very useful concept for two main reasons. First, it is redundant with the item difficulty index which is the proportion of students answering the item correctly (see Brown, 1976; French, 2001). Note that a lower proportion reflects a higher item difficulty. The total item distraction index is then simply *one minus* the item difficulty index. Of course, item difficulty could also be defined as "proportion correct of high scorers plus proportion correct of low scorers altogether divided by two" (see Brown, 1976, p. 270).

Second, consider two multiple-choice test items each with three decoys or distractors. Assume the first test item has the following individual distraction indexes--0.10, 0.15, 0.20, and the second test item has the following individual distraction indexes --0.00, 0.00, 0.45. The two test items will have the same total item distraction index of 0.45. Yet, intuition tells us that the distractibility of the latter test item's three decoys collectively is NOT as high as the former test item's three decoys. A proper *collective distractibility* (of three decoys) index should reflect a higher value for 0.10, 0.15, and 0.20 than for 0.00, 0.00, and 0.45. Applying the present paper's concept of choice-agreement index would solve this problem. We could define the *collective distractibility* index of K decoys of a test item as the present paper's *choice-agreement index*. Note that, by equating in this manner, a lower value of *collective distractibility index* reflects a higher degree of distractibility. This is in line the classical notion of item difficulty index mentioned earlier in which a low value also indicates higher level of difficulty.

It is important to note that whereas the *total item distraction index* (mentioned earlier) is totally redundant with the item difficulty index and therefore yielding a perfect correlation between the two indexes (variables), the *collective distractibility index* should NOT correlate perfectly with the item difficulty index. It is easy to see that a test item's difficulty is not totally dependent on its decoys' collective distractibility by the following example. Consider two standard four-choice test items with proportions of choosers for the first item as 0.70, 0.10, 0.10, 0.10, and for the second item as 0.40, 0.20, 0.20, 0.20. Let the first choice in both items be the correct choice. It is then clear that while the difficulty of the two items is different (0.70 vs. 0.40), the *collective distractibility* is the same (0.00). The actual correlation between the *item difficulty index* and the *collective distractibility index* can only be determined empirically in an actual test situation.

As an exploratory investigation of the correlation between *item difficulty index and collective distractibility index*, the author has used the data from a final exam given in a research course offered last summer. There were 25 students and 50 multiple-choice test items in the final exam. Thirty-nine of the 50 items had the standard three decoys while the other 11 items had either one, two or four decoys. Only the 39 items with the standard three decoys were retained for possible analysis. Of these 39 items, six items were answered correctly by every student and therefore eliminated from analysis--leaving 33 items. Of these 33 items, six items had only one student choosing a decoy and were therefore also eliminated from analysis--leaving 27 items for final analysis. Note that our *collective distractibility index* is based on the *choice-agreement index* which is based on the notion of agreement pairs and disagreement pairs of choosers. Therefore, we need at least two students choosing the decoy(s). The final 27 test items each with its computed *item difficulty index*--defined as the proportion of students answering an item correctly (see Brown, 1976; French, 2001), and its computed *collective distractibility index*--defined in the present paper, are as shown in Table 3. The correlation coefficient between the two indexes is 0.49, $p < .01$.

Summary and Conclusion

Past item analysis of distractors of a test item tended to focus on the property of each distractor or decoy separately. Green, Crone and Folk (1989), for example, studied the attractiveness of each different distractor to test takers of different ethnic backgrounds and ability levels (Differential Distractor Functioning analysis). Thissen, Steinberg, and Fitzpatrick (1989) presented an item response model which analyzes the probability of choosing the correct alternative and each of the distractors of a test item as a function of a test taker's proficiency. The present paper, with the concept of collective distractibility index, provides a means to focus on the collective property of all distractors of a test item. The example analysis in the previous paragraph has already established a possible significant correlation between the collective distractibility of an item and the item's difficulty. Future studies might investigate the differential collective distractibility of a test item among different groups of test takers.

Application of the Choice-Agreement Index presented in the present paper could be applied to other areas than education. In economics, consumers have to choose among different brands while shopping. Companies may try to measure choice-agreement among consumers and try to change it through advertisement campaign. The pre-post Choice-Agreement Indexes computed and the significant difference between them could serve to determine the success of the advertisement. In politics, voters have to choose among different candidates. Candidates may try to measure choice-agreement among voters and try to change it through a political campaign. From before and after campaign polling, the pre-post Choice-Agreement Indexes computed and the significant difference between them could serve to determine the success of the political campaign. Application of the Choice-Agreement Index in other areas than education, economics, and politics should be possible.

References

Anastasi, A. (1982). *Psychological Testing* (5th ed.). New York: Macmillan.

Brown, F. G. (1976). *Principles of educational and psychological testing* (2nd ed.). New York: Holt, Rinehart, and Winston.

Epstein, M. H., Cullinan, D., Harniss, M. K., & Ryser, G. (1999). The Scale for Assessing Emotional Disturbance: Test-retest And interrater reliability. *Behavioral Disorders,* 24(3), 222-230.

Fan, X., & Chen, M. (2000). Published studies of interrater reliability often overestimate reliability: Computing the correct coefficient. *Educational and Psychological Measurement,* 60(4), 532-542.

French, C. L. (2001). *A review of classical methods of item analysis.* Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 1-3, 2001. (ERIC Document Reproduction Service No. ED 450-152)

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45,* 135-145.

Green, F. F., Crone, C. R., & Folk, V. G. (1989). A method for studying differential distractor functioning. *Journal of Educational Measurement,* 26(2), 147-160.

Hays, W. (1973). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart and Winston.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6(3),* 153-160.

Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin, 76(5),* 365-377.

Light, R. J. (1973). Issues in the analysis of qualitative data. In R. M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 318-381). Chicago: Rand McNally College Publishing.

Kendall, M. G. (1948). *Rank correlation methods.* London: Griffin.

Robinson, W. S. (1957). The statistical measurement of agreement. *American Sociological Review,* 22, 17-25.

Rosner, B. (2000). *Fundamentals of biostatistics* (5th ed.). Pacific Grove, CA: Duxbury.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15,* 72-101.

TenBrink, T. D. (1974). *Evaluation: A practical guide for teachers.* New York: McGraw-Hill.

Thissen, D., Steinberg, L., & Fitzpatrick, A. R. (1989). Multiple-choice model: The distractors are also part of the item. *Journal of Educational Measurement, 26*(2), 161-176.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.), International Student Edition. Tokyo, Japan: McGraw-Hill Kogakusha.

Appendix
Maximum Number of Disagreement Pairs

Given N choosers and K alternatives, find the frequency distribution that constitutes the maximum number of disagreement pairs.

Let n_1 be those who choose alternative 1, n_2 alternative 2, ... , n_K alternative K. Then, the number of disagreement pairs is $n_1n_2 + n_1n_3 + \dots + n_1n_K + n_2n_3 + n_2n_4 + \dots + n_2n_K + \dots + n_{K-1}n_K$ or, in short, $\sum\sum_{i<j} n_in_j$ (where i = 1, K-1; j = 2, K)

To find the relative maximum or minimum of $\sum\sum_{i<j} n_in_j$, first partially differentiate $\sum\sum_{i<j} n_in_j$ with respect to $n_1, n_2, \dots n_K$ and then set the results equal to 0 and solve for $n_1, n_2, \dots n_K$.

Let $f = \sum\sum_{i<j} n_in_j$.

Since $(n_1 + n_2 + \dots + n_K)^2 = n_1^2 + n_2^2 + \dots + n_K^2 + 2\sum\sum_{i<j} n_in_j = N^2$,

$$\sum\sum_{i<j} n_in_j = \frac{1}{2}[N^2 - (n_1^2 + n_2^2 + \dots + n_{K-1}^2 + n_K^2)]$$

$$= \frac{1}{2}[N^2 - n_1^2 - n_2^2 - \dots - n_{K-1}^2 - (N - n_1 - n_2 - \dots - n_{K-1})^2]$$

since $n_K = N - n_1 - n_2 - \dots - n_{K-1}$.

Therefore, $f = \frac{1}{2}[N^2 - n_1^2 - n_2^2 - \dots - n_{K-1}^2 - (N - n_1 - n_2 - \dots - n_{K-1})^2]$.

Noting that N is a constant and n_i's are variables, we perform partial differentiation:

$$\frac{\partial f}{\partial n_1} = -n_1 + (N - n_1 - n_2 - \dots - n_{K-1})$$

$$\frac{\partial f}{\partial n_2} = -n_1 + (N - n_1 - n_2 - \dots - n_{K-1})$$

. .

. .

. .

$$\frac{\partial f}{\partial n_{K-1}} = -n_{K-1} + (N - n_1 - n_2 - \dots - n_{K-1})$$

Setting the results equal to zero, we have

$$-n_1 + (N - n_1 - n_2 - \ldots - n_{K-1}) = 0 \quad \text{(A1)}$$

$$-n_2 + (N - n_1 - n_2 - \ldots - n_{K-1}) = 0 \quad \text{(A2)}$$

$$\vdots$$

$$-n_{K-1} + (N - n_1 - n_2 - \ldots - n_{K-1}) = 0 \quad \text{(A[K-1])}$$

It is clear that, in the above set of equations, $n_1 = n_2 = \ldots = n_{K-1}$ since every one of them is equal to $(N - n_1 - n_2 - \ldots - n_{K-1})$. Noting that the term in parentheses is in fact n_K, we deduce further that $n_1, n_2, \ldots, n_{K-1}$ are all equal to n_K and therefore $n_1, n_2, \ldots, n_{K-1}, n_K$ are all equal to one another.

Since $n_1 + n_2 + \ldots + n_K = N$ and $n_1 = n_2 = \ldots = n_K$, we must have $n_1 = n_2 = \ldots = n_K = N/K$.

Since $\frac{\partial^2 f}{\partial n_i^2}$'s (i = 1, K-1) are all negative, we have the situation of a maximum.

Therefore, the maximum of $\sum\sum_{i<j} n_i n_j$ or the maximum number of disagreement pairs occurs when $n_1 = n_2 = \ldots = n_K = N/K$, and the total number of maximum disagreement pairs is:

$$\binom{K}{2} \frac{N}{K} \cdot \frac{N}{K}$$

$$= \frac{K(K-1)}{2} \cdot \frac{N^2}{K^2}$$

$$= \frac{K(K-1)N^2}{2K} .$$

Table 1

Distribution of Disability Income Insurance Choices Made by 194 College Personnel

Choice	N	%
TIAA	22	11
NSEA	159	82
Undecided	13	7
Total	194	100

Table 2

Level of Agreement and Patterns of Choices with Two Alternatives and Six Choosers

Level of Agreement	One alterntive (# people)	The other alternative* (# of people)	$Index_2$
1	0	6	1.00
2	1	5	0.44
3	2	4	0.11
4	3	3	0.00

Note. * The order of alternatives does not matter. If 1 person chooses alternative A and 5 choose alternative B, the level is considered to be the same as when 1 person chooses B and 5 choose A.

Table 3

Item Difficulty Index and Collective Distractibility Index of 27 Test Items

Item #	Item difficulty index	Collective distractibility index
1	0.84	1.00
2	0.88	1.00
3	0.76	1.00
5	0.64	0.19
8	0.56	0.18
10	0.88	0.58
13	0.88	1.00
14	0.92	0.50
15	0.40	0.61
17	0.48	0.20
19	0.88	1.00
20	0.88	0.58
21	0.56	0.51
23	0.92	1.00
25	0.84	1.00
26	0.76	1.00
27	0.88	0.58
29	0.76	0.29
32	0.84	0.50
34	0.88	1.00
35	0.64	0.38
36	0.80	1.00
41	0.80	0.20
43	0.88	1.00
45	0.76	1.00
46	0.60	0.72
50	0.88	0.58