The Mathematics Enthusiast

Volume 12
Number 1 *Numbers 1, 2, & 3*

Article 27

6-2015

Judgment of Association between Potential Factors and Associated Risk in 2x2 Tables: A Study with Psychology Students

Carmen Batanero

Gustavo R. Cañadas

Carmen Díaz

Maria M. Gea

Judgment of Association between Potential Factors and Associated Risk in 2x2 Tables: A Study with Psychology Students

Carmen Batanero
University of Granada, Spain
Gustavo R. Cañadas
University of Granada, Spain
Carmen Díaz
University of Granada, Spain
Maria M. Gea
University of Granada, Spain

Abstract: This study was aimed to evaluate the accuracy and strategies used in the estimation of association between potential factors and associated risks when data are presented in 2x2 tables. A sample of 414 undergraduate Psychology students from three different Spanish universities was given three different tasks (direct and inverse association and perfect independence) where they had to estimate such association. Most participants judged association in the task where there was perfect independence, but the data contradicted the students' previous expectations. The estimation of association was consistent with the perception of association and the accuracy of estimates increased with correct strategies. Our participants performed worse than secondary school students in a previous study (Batanero et al., 1996) and we found no difference in the three participating universities. We classify the students' strategies in the tasks in levels of complexity and explain the incorrect strategies using the idea of semiotic conflict.

Keywords: risk factors, association, 2-way tables, semiotic conflicts, psychology students.

Introduction

In the past decades we have observed an increasing interest towards risk perception and management in many organizations, such as private companies, banks, hospitals, or schools. Slovic (2000) described a risk as an uncertain event or condition that when happening has a positive or negative effect on a person or a group of people. Gigerenzer (2003) suggested that such an event should be associated with a probability or a frequency with is based on empirical data about its past or potential occurrence. In this definition the author includes different interpretations of probability; such as subjective probability (degree of belief of the person assigning the probability), propensity (physical property of an object; for example, regularity in a dice) or frequentist estimation (information from a large number of observations of the event). Consequently, an event could be perceived or not as a risk, depending on the person's conception of probability.

Today an increasing number of events are described in terms of risk using mathematical formats, such as probabilities, proportions or percentages. The underlying concepts have to be learned in school and for this reason, mathematics educators are becoming interested in students' perception and understanding of risk (Martignon, 2014). In the same way, the assessment of students' potential biases, wrong strategies and misconceptions when interpreting risks is a relevant area of research in mathematics education (Nunes, & Bryant, 2011).

In health or clinical contexts, including psychological evaluation or diagnosis, risk is synonymous of hazards and dangers; for example an illness or the undesirable effect of a treatment (Power, 2007). Risks in these contexts are often associated to decision making in such a way that it is

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 347-363

impossible to make a 'risk-free' decision, unless we leave some potential risk factors unmanaged. A sound understanding of risk and of the associated numerical information is essential for these professionals, in order to make adequate decisions. However, numerous examples that this understanding is not complete are described in the literature related with risk perception and management. For example, Gigerenzer et al. (2007) termed *collective statistical illiteracy* the fact that many professionals have difficulties in interpreting health statistics and draw wrong conclusions in the clinical practice without noticing. Gigerenzer and Edwards (2003) suggest that misperception of risk factors occur by confusing single event probabilities, conditional probabilities, and relative risks. They also suggest that the situation can be improved by education and by representing the information regarding risks in ways that are transparent for the human mind, such as natural frequencies, tree diagrams or two-way tables.

In fact, two-way tables are a main representational tool for bivariate data, and are often used in professional journals to report the influence of risk factors on different pathologies. Its understanding is related to risk perception (Nunes, & Bryant, 2011). Adi, Karplus, Lawson, and Pulos (1978) suggested that providing subjects with information already organised in tables improves their performance in tasks in which they are asked to assess whether there is an association between two events.

These tables; in particular 2x2 tables (two-way tables with only two columns and two rows) are also an important tool in diagnosis and psychological evaluation, where psychologists are confronted with different potential risk factors that may be associated or not with a disorder (risk) (Diaz, & Gallego, 2006). The estimation of association in these tables is the first step to determine which factors are associated to specific risks. However, even when association judgment is a priority learning issue in statistics courses (Zieffler, 2006), little attention is paid to its teaching, in assuming that the interpretation of 2x2 tables is easy. Contrary to this belief, previous research on this area described in Section 2 suggests that people's performance in judging association from data presented in a 2x2 tables is often inaccurate.

This study was aimed to evaluate the accuracy and the strategies used in the estimation of association in 2x2 tables by psychology undergraduate students. A sample of 414 students of Psychology in three different Spanish universities was given three different tasks (where data show direct and inverse association and perfect independence). We compare the judgment of association and the accuracy in the estimation of the strength of association by item and university. The strategies used by the students to judge the existence of association are classified in five levels of complexity described by Pérez-Echeverría (1990). Our results are compared with a previous study with high school students by Batanero, Estepa, Godino and Green (1996). We use the idea of *semiotic conflict* introduced by Font, Godino and D'Amore (2007), to explain the prevalence of incorrect strategies. Some final implications for teaching 2x2 tables are included.

Background

Perception of Risk

The interest in analyzing people's perception of risk arose by the observed fact that people often disagree about the likelihood that a risk happens in presence of some risk factors. Several theories have been proposed to explain why different people make different estimates of the likelihood of risks. Early work in Psychology (e.g., Tversky, & Kahneman, 1974) suggested that people use cognitive *heuristics* in sorting and simplifying information in uncertain situations which lead to biases in decision making. This research also suggested possible explanation of peoples' biases in decision making in terms of heuristics, such as *representativeness, availability or anchoring,* some of which may affect people's perception of risk. For example, the representativeness heuristics occurs when people judge the probability of an event by considering only its similarity with a population. An associated bias is the *insensitivity to prior probabilities,* where people do not take into account a condition that affects the probability for such event. In another example, the *illusion of validity* is the belief that irrelevant information (e.g., personal experience, salience of information) generates additional relevant data for predictions, even when this is not the case.

Association in 2x2 Tables

Between the complete certainty of some events and the complete uncertainty lies the world of association. Association between some variables (factors) and uncertain events (risks) are frequent in many situations both in professional and everyday life. Correct perceptions of these associations help people evaluate the probability of a particular risks taking place in presence of an associated factor.

Correlational reasoning is an important research area which is aimed to describe people's accuracy and strategies to evaluate the presence and strength of association between variables (see Adi, Karplus, Lawson, & Pulos, 1978, for a survey). According to this research, correlational reasoning is an important component of scientific reasoning, as it is a tool for understanding the past, controlling the present and predicting the future (Alloy & Tabachnik, 1984).

Research on correlational reasoning has been often carried out with dichotomous variables; where we use the word *associatio*n to describe the relationship between these variables (the word correlation is used to describe relationships between numerical variables). In this research participants are given some tasks where they are asked to decide if a variable with two modalities (*A,* not *A*) is associated or not with another variable with two modalities (*B,* not *B*).

The pioneer work in this line of research was due to Inhelder and Piaget (1955), who conceived association as the last step in the development of probabilistic reasoning. They performed interviews with 13- 14 year olds students and proposed them the study of the association between eyes colour (blue and dark eyes) and hair colour (fair and brown hair) using data that were formally equivalent to Table 1. According to Nunes and Bryant (2011), understanding the association between the two variables in Table 1 makes three demands on people's reasoning: a) they need to understand that the situation involves randomness, since not always *A* and *B* appear together; b) it is important to recognise the cases (*a* and *d*) that support an association between the variables and the cases (*b* and *c*) that go against the association; and c) the subjects should be able to quantify and compare the positive and negative cases in order to assess whether the co-occurrence observed is not due to chance.

	A	Not *A*	Total
B	a	b	a+b
Not *B*	c	d	c+d
Total	a+c	b+d	a+b+c+d

Table 1. Scheme of a 2x2 table.

Inhelder and Piaget (1955) described different types of strategies used by the children in their interviews to decide if there is association between the variables or not. In a first stage (13 year olds or younger children) the subjects only analyze the favourable positive cases in the association (cell *a* in Table 1). In a second stage of development they compare *a* with *b* or compare *c* with *d.* Inhelder and Piaget suggested that, although these children can compute single probabilities, understanding association requires considering quantities *(a+d)* as favourable to the association and *(b+c)* as opposed to association, and also that it is necessary to consider the relation:

$$R = \frac{(a+d)-(b+c)}{a+b+c+d}$$

where *R* represents the difference between cases confirming and opposed to the association compared to all the possibilities. We can observe that *R*=1, when all the data fall in cells *a* and *d*; *R*=-1 when all the data fall in cells *b* and *c* and *R*=0 when *(a+d)= (b+c).* According to Piaget and Inhelder, recognition of this fact only happens at 15 years of age.

The study by Inhelder and Piaget inspired some further research, and was extensively repeated with subjects at different ages and this research showed that adults have poor correlational reasoning. For example, Smedslund (1963) found adults who based their judgment only on the frequency in cell *a* in Table 1 or by comparing the frequencies in cells *a* and *b.* Allan and Jenkins (1983) showed the tendency to base the association judgments on the difference between the frequency of cell *a* (cases where *A* and *B* happen simultaneously) and frequency in cell *d* (where neither *A* or *B* happen). They remarked that these subjects do not understand that a high frequency in cell *d* is favourable to a positive association. Jenkins and Ward (1965) pointed out that even the strategy of comparing the probabilities of the table diagonals *(a+d)* and *(c+b)* (computing *R*) considered correct by Piaget and Inhelder is not always valid; if the difference in the totals in the rows or columns in the table is high, this strategy can produce errors. Nevertheless, in Allan and Jenkins's (1983) research this strategy was widely used by adults.

According to Jenkins and Ward the correct general strategy requires comparing conditional probabilities; for example, comparing the probability of *B* in the group of people with *A* and in the group of people with not *A:*

$$D = P(B \mid A) - P(B \mid notA) = (\frac{a}{a+c}) - (\frac{b}{d+d})$$

Pérez-Echevarría (1990) classified all the strategies that have been identified in judging association in 2x2 tables in 5 different levels of complexity: In levels 0 to 3 people only use data in 0 to 3 different cells; in Level 4 the subjects base their judgment of association on additive comparisons of the data in the four cells. In Level 5 the subjects base their judgment of association on multiplicative comparisons between the four cells. These level 5 strategies are the only correct strategies in general; although in some particular tables the subjects can succeed with lower level strategies.

Variables that Affect Perception of Association

Other researchers have focussed on the evaluation of people's accuracy when they estimate the strength of association in 2x2 tables. A typical task in this research is asking people to provide a number between 0 (perfect independence) and 1 (perfect association), according to their perception of their strength of association in the table data. The number provided by the subjects is compared with a statistical index of association in the data (for example, with the Pearson's Phi coefficient).

The main aim of this research is to find the variables that affect the accuracy in the estimation. Crocker (1981) showed that this accuracy is higher when the cell frequencies are smaller; and also when the events co-vary simultaneously through time; for example, when we consider hair and eye colour of the same subjects. Erlick and Mills (1967) found that negative association is estimated as close to zero. Arkes and Harkness (1983) indicates that the frequency in cell *a* has the greatest impact on the accuracy of estimates. The accuracy of estimation is higher in causal contexts. According to Barbancho (1992), an association between variables may be explained by the existence of a unilateral cause-effect relationship (causal context). However we may find association in the case of interdependence (each variable affects the other), indirect dependence (there is a third variable affecting the other two) or spurious covariation (the association happens by chance). In addition to the estimate accuracy, understanding association involve the discrimination of these types of relationships between variables.

Biases and misconceptions in judging association

Judging association is also subjected to cognitive biases: *Illusory correlation* takes place when people perceive a relationship between variables that do not exists, stronger than existing of opposite to the data association:

> The report of a correlation between two classes of events which, in reality, (a) are not correlated, (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction from that which is reported (Chapman, 1967, pp. 151).

We could consider illusory correlation as a particular instance of the illusion of validity described by Tversky and Kahneman (1974). Many researchers have found that illusory correlation is common and influences the estimates of association (Jennings, Amabile, & Ross, 1982; Wright & Murphy, 1984; Meiser & Hewstone, 2006). The estimates of association are more accurate if people have no prior expectations about the type of association in the data. If the subject's prior expectations agree with the type of association reflected by the empirical data, there is a tendency to overestimate the association coefficient. But when the data do not reflect the results expected by these personal theories, the subjects are often guided by their beliefs, rather than by the data.

Other biases have been found in previous research. Batanero, Estepa, Godino, & Green (1996) analyzed the strategies in association judgments by 213 high school students (16-17 year-olds) and defined different conceptions of association: (a) *causal conception* according to which the subject only considers association between variables, when it can be explained by the presence of a cause - effect relationship; (b) *unidirectional conception*, where students does not accept an inverse association, (c) *local conception,* where the association is deduced from only a part of the data, and determinist conception, when the student considers association only in case of functional (perfect) dependence.

Mathematical Objects Linked to Two-Way Tables and Potential Semiotic Conflicts

Although previous research related to judgment of association in 2x2 tables is abundant, none of this research questioned the students' capacity to interpret the different mathematical objects implicit in the analysis of Table 1, which is a complex semiotic object (Estrada, & Díaz, 2006). Data in cells *a, b, c, d* refer to joint absolute frequencies, each of them for a double condition (values of row and column); however; as suggested by Inhelder and Piaget (1955), their meaning is non-equivalent. A high frequency in cells *a* (presence of character A; presence of character B) and *d* (absence of A, absence of B) would indicate a positive association between the variables; and a high frequency in the other two cells suggests a negative association. Moreover, from a given cell we can deduce joint as well as row and column conditional relative frequencies. For example, from cell *a,* we can compute three different relative frequencies (or three different percents if we multiply each of these relative frequencies by 100):

- Joint relative frequency: $\frac{a}{a+b+c+d}$
- Relative frequency as regards the row total: $\frac{a}{a+b}$
- Relative frequency as regards the column total: $\frac{a}{a+c}$
- Furthermore, we can compute the relative marginal frequencies of rows and columns: $\frac{a+b}{a+b+c+d}$ and $\frac{a+c}{a+b+c+d}$

All these mathematics objects co-exist and may be confused by students. In our analysis we use the construct *semiotic conflict* taken from Font, Godino and D'Amore (2007). These authors adapted from Eco (1979) the idea of *semiotic function* or correspondence between an expression and its content and suggest that, in mathematical practices many objects intervene: problems, actions, concepts-definition, language properties and arguments, any of which could be used as either expression or content in a semiotic function. The authors termed *semiotic conflict* any disparity or difference of interpretation between the meanings ascribed to an expression by two subjects.

Consequently, we assume that some incorrect strategies used in judging association maybe explained by misinterpretation or confusion between the different mathematical objects involved in the analysis of 2x2 tables. For example Falk's (1986) described the *fallacy of transposed conditional* by which students do not adequately discriminate the two different conditional probability $P(A|B)$ and $P(B|A)$. Einhorn and Hogarth (1986) observed that some students misinterpret the conjunction "and" when computing probabilities and confuse joint and conditional probability. It is possible that these errors appear when interpreting a 2x2 table and this led the students to use an incorrect strategy to judge association.

Wiht this conjecture in mind, our research was aimed to assess the accuracy in the estimation of association in 2x2 tables by Psychology students. We also try to complement the analyses of strategies made by Batanero et al. (1996) by identifying students' semiotic conflicts when judging association.

Method

The sample included 414 students in their first year of Psychology studies from three Spanish universities: Almeria (115 students), Granada (237 students) and Huelva (62 students), all of them taking an introductory statistics course. The questionnaire was given to the student as a part of a practice. The samples included all the students enrolled in the course and attending the session; the difference in sample sizes was due to the size of the University: Almeria with 2 groups of students, Huelva with 1 group of students and Granada with 4 groups of students. Though the students had not yet studied association in the course they were following; however, they had studied descriptive statistics and probability in the first semester.

The questionnaire was adapted from Batanero, Estepa, Godino and Green (1996) (see Appendix). The context was changed to psychological diagnosis in two items (1 and 2). The frequencies in the table cells were increased in items 2 and 3, since in the original questionnaire the small sizes made the application of the Chi-square statistics invalid. The sign and strength of association were maintained in all the items. The following task variables (Table 2) were considered in the questionnaire:

1. *Sign of association:* We include the three possible cases: direct and inverse association and independence.

2. *Strength of association,* which was measured by the Pearson's Phi coefficient. An item with perfect independence and two items with moderate-high association were included.

3. *Agreement between association in the data and previous theories* suggested by the context. There was one item where the empirical association matched the prior expectations and one where it contradicted the expectations and another with a neutral context suggesting no previous theories.

4. *Type of covariation.* We used three categories of Barbancho's (1992) classification: unilateral causal dependence, interdependence and indirect dependence.

	Item 1	Item 2	Item 3
Sign of association	Independence	Inverse	Direct
Association coefficient (Pearson's Phi)	0	-0.62	0.67
Agreement with prior theories	No	Yes	There is no theory
Type of covariation	Interdependence	Causal unilateral	Indirect dependence

Table 2. Task variables in the items

A qualitative analysis of students' responses served to define three different variables:

1. In part (a) of each item, students are asked to provide an association judgment. We considered three different categories in their responses: (a) the student considers that the variables in the item are related (association); (b) the student considers that the variables are not related (independence); and (c) the student does not decide (no judgment).

2. The estimation of the strength of association is deduced measuring the exact position of the point drawn by the student on the numerical scale (0-1) in the second part of the item.

3. Finally, a qualitative analysis of the reasoning given by the students to justify their responses serves to identify the strategies used by the students and their semiotic conflicts. The classification of strategies was performed independently by two different members of the team; in the case of disagreement, it was revised by other team members until an agreement was reached.

Perception of Association

Association Judgment

In order to analyse the students' competence to perceive the possible association between the data given in each item, we present the percentage of students who accepted the existence of a relationship between the variables in Table 3. In the last columns we add the value of the Phi association coefficient for the data and the relationships between prior theories and data.

	Almería (n=115)	Granada (n=237)	Huelva (n=62)	Total (n=414)	Association coefficient	Prior theories vs data
Item 1	87 (75.7)	194 (81.9)	42 (67.7)	323 (78.1)	0	Do not agree
Item 2	108 (93.9)	230 (97.1)	60 (96.8)	398 (96.1)	-0.62	Agree
Item 3	107 (93.1)	226 (95.4)	53 (85.5)	386 (93.2)	0.67	No theories

Table 3: Frequency (and percent) of students considering there is association in the data.

Most students indicated the existence of association in all items, in particular when the association was confirmed by the data, but also in Item 1 (where the data correspond to perfect independence). We explain this result by the mechanism of illusory correlation (Chapman, 1967), since in this item data contradicts the students' previous expectation (that stress is related to insomnia). Many students expressed their previous belief that both variables should be related in this item in their answers: "*There should be some relationship, since in my experience stress due to family or other type of problems may be a cause of insomnia*" (Student 213). *"In my opinion insomnia and stress are related, since most people who have insomnia suffer from stress"* (Student 5), or "*Yes, because people with insomnia do not rest well and this causes extra stress that is added to stress due to other external factors*" (Student 42).

These responses and the high percentage of students that argued that the variables are associated (in spite of perfect independence in the data) suggest that these previous theories affect the student perception of association. Moreover these responses show that part of these students might identify correlation and causation. Our results in this item are worse than those in Batanero et al. (1996) with secondary school students, since these authors only found 55.4% of students judging association in an item with exactly the same data and a similar context (smoking and bronchial disease).

Our students outperformed in Item 2 (inverse association) those in Batanero et al. (2006), where only 47.1% of students considered association. This result could be explained by the change in context (diet and digestive troubles in Batanero et al.) and the increased sample size in our item. Consequently, the unidirectional conceptions of association described by Batanero et al. (1996) by which students perceive negative association as close to independence hardly appear in our research. Results in Item 3 were very close in both studies (92.5% in Batanero et al., 1996), where we only increased the frequencies without changing the context or the strength of association.

In Table 3 we observe little differences in the percentages of students accepting association in each item in the three universities. In order to check the homogeneity of these samples we carried out a Chi-square test of homogeneity between the three samples and we obtained the following results (Chi= 0.99; 6 d.f.; p=0.9861). This result led us consider the subsamples to be homogeneous as regards the percentage of students considering association in each item and that there was no much difference in this variable, in spite of the different educational context.

Estimating the strength of association

In the second part of their response to each item, the students provided a score between 0 and 1 according to the association strength they perceived in the data. This score is an estimate of the association coefficient (disregarding the sign). In Table 4 we present the mean score obtained by university and in the whole sample by item, as well as the characteristics of each item. The most accurate estimate corresponds to Item 3, where the students showed no prior expectations: The average estimate is very close to the empirical coefficient in all the samples and in the whole sample.

	Mean estimate				Association coefficient	Prior theories vs data
Item	Almeria (n=115)	Granada (n=237)	Huelva (n=62)	Total (n=414)		
1	0.51	0.47	0.44	0.47	0	Do not agree
2	0.78	0.72	0.73	0.73	-0.62	Agree
3	0.75	0.68	0.68	0.70	0.67	No theories

Table 4. Mean estimates of the association coefficient

In Item 1, which corresponds to perfect independence, the global mean value was 0.47, and about the same value in each sample. As we described in the previous section, the students' previous expectations contradicted the independence in the data in this item. Moreover, in this item cell *a,* (simultaneous presence of both stress and insomnia) contains the maximum absolute frequency; according to Arkes and Harkness (1983) cell *a* has the greatest impact on the subject's attention,.

The estimate for Item 2 (inverse dependence), was higher than the empirical value association in all universities. Thus, in our students we did not find a significant presence of the unidirectional conception described by Batanero et al. (1996). In this item the students' previous expectations coincided with the association in the data. Some students expressed these beliefs in their response: "*I think there is relationship between having brothers or sisters and being a problematic child, since children having brothers or sisters are raised with different moral values, such as generosity or empathy, than only children*" (Student 44); "I*n my experience only children are troublesome"* (Student 172).

Moreover, both in the whole sample and in each university most students indicated that there was association in this item (the sign of association was not requested).

This judgment of association was consistent with the students' estimation of the association coefficient (Table 5) with the differences in the mean estimates in students judging association in the item or not judging association being statistically significant in the t-test of differences for all the items.

	Association	No association				
Item	Mean estimate	Mean estimate	Mean difference	Standard error for the mean difference	Students t	p-value
1	0.56	0.18	0.38	0.02	17.00	0.000
2	0.75	0.31	0.44	0.05	8.76	0.000
3	0.73	0.22	0.51	0.04	12.10	0.000

Table 5. Mean estimates of the association coefficient in students considering or not association in the data and t- test of differences in means

The students from Granada and Huelva estimated an average lower association in all the items than the students from Almeria; however, the differences in mean estimates were not statistically significant in the one-way ANOVA tests (F=0.487, p=0,615 for Item 1; F=0.260, p=0,771 for Item 2 (F=0.308, p=0.735 for Item 3 with d.f=2,411). These results suggest that we cannot reject the hypothesis that the students' average estimates are similar in the three universities, despite the possible differences in educational context. Consequently we decided to combine the three samples for further analyses.

Students' Strategies

Level of strategies

To complement the above results we performed a qualitative analysis of students' strategies. Firstly, we differentiated three groups of strategies:

Correct strategies. These always produce a correct association judgement, such as comparing conditional frequencies by rows or columns. An example is comparing the proportion of people with stress disorder with those with no stress disorder in people with and without insomnia in Item 1.

Partly correct strategies. Procedures that provide a correct association judgment for the particular table in which it is applied, but that is not valid in general for any 2x2 table. An example in Item 2 is comparing the frequencies in the first row; as we found less problematic children among those who have sisters or brothers, the student correctly deduces association in the data. The procedure works for this example, but not for the general case; for example, however, this procedure leads to an incorrect judgment in Item 1.

Incorrect strategies. When students use a procedure that is incorrect in all type of tables. A particular example is not using the data table; another example is using only one cell, as in this case it is impossible to study the variation of frequencies with different combinations of the variables.

This classification was crossed with the levels of difficulty proposed by Pérez-Echeverría (1990). All level 0 and 1 strategies are incorrect, while part of the strategies in upper levels are partly correct or correct.

Level 0 Strategy. The student uses no data from the table and only takes into account his/her own previous beliefs about the association that should be in the variables; the illusory correlation (Chapman & Chapman, 1969) is visible in these students; for example: "*The variables are related, since when you do not sleep your stress increases*" (Student 5).

Level 1 Strategy. When the student only used one cell in the table. Usually the student use only cell *a* because in this cell when both characters are present and so the frequencies in this cell have a higher impact on our attention (Smedlund, 1963; Beyth -Marom, 1982; Shaklee & Mins, 1982, Yates & Curley, 1986): *"there is association, since 90 people (36 % of the sample have insomnia and stress"* (Strategy 1.1; Student 111). Other students used one of the cells *b* or *c* that contradicts the association: *"There is no relation since there are 60 people with stress and no* *insomnia and this is a big percentage"* (Strategy 1.2, Student 51). These students expect a deterministic relation between the variables as they assume that there could be no exception in the association; that is, he expected that all people with stress should suffer from insomnia.

Level 2 Strategy. Some students use two cells; for example, they compare *a* with *b* or *a* with *c*. Consequently they deduce the association from only one conditional distribution in the table: *"If you look to the people with insomnia, there are more people with stress (90) than without stress (60)"* (Strategy 2.1, Student 21). In this item this strategy is incorrect, as it led to an incorrect judgment; however, if used in items 2 or 3 is partly correct as it will produce a correct judgment. Other students compared the cells with maximum and minimum frequency: *"There are 90 people with stress and insomnia and 40 without stress and without insomnia; 90>40, but the relation is not too strong"* (Strategy 2.2, Student 61). This strategy is incorrect, because cell *d* is also relevant for the association, contrary to what the student assumes.

Level 3 Strategy. In this strategy the student uses three cells; for example, he compares cell *a* with *b* and *c*. In general, these students discarded cell *d* that corresponds to the absence of both characters: *"There is relationship as there are more people with stress and insomnia (90 people) and exactly the same number (60) with either stress and no insomnia or insomnia and no stress"* (Strategy 3.1, Student 153). The strategy could work with items 2 and 3 and then is partly correct for these items but it is incorrect for Item 1.

Level 4 strategies. These strategies are based on additive comparisons of the four cells. One example is comparing the sum of diagonals *(a+d)* with *(b+c)*: *"There are 130 people with both stress and insomnia or no stress and no insomnia, while there are only 120 with one of these symptoms.* (Strategy 4.1, Student 176). This strategy was found by Allan and Jenkins (1983) and could provide a good solution when the marginal frequencies (number of people with and without insomnia) were equal, according to Shaklee (1983. As we see in the example, it does not work with Item 1; for this reason we consider this strategy to be incorrect for this item. In another example, students compared two conditional distributions in an additive way: *In people with insomnia there is a difference of 30 having stress, while the difference in people without insomnia is smaller (20)"* (Strategy 4.2, Student 267). Finally other students compared all the absolute frequencies among them: "*There are many people with stress and insomnia (90) but the relationships is not strong, since having stress and no insomnia or insomnia and no stress (60) is also high, much higher than no insomnia and no stress (40)* (Strategy 4.3, Student 156).

Level 5 strategies. Some students use all the four cells with multiplicative comparisons, but still may be incorrect or partly correct. For example, a wrong strategy is to compute all the joint relative frequencies and compared them: *"I computed the percent of each data and compared the results:* $\frac{90}{250}100 = 36\%$; $\frac{60}{250}100 = 24\%$; $\frac{60}{250}100 = 24\%$; $\frac{40}{250}100 = 16\%$ "(Strategy 5.1, Student 11). This procedure is incorrect, because the association should be deduced from conditional distributions and not from joint distributions. An example of partly correct strategy is assuming that all joint relative frequencies in the table should be identical (that is, 25% cases in each cell). We considered this strategy partly correct because the student computed some "expected" frequencies, compared them with the observed frequencies and correctly deduced that there was association because these two types of frequencies were different. However the strategy is not valid in the general case.

> I divided 250 between 4 (25%) to see the number of children we should expect in each cell, in case of no relationship. However, although the number of only child who are problematic are close to 25% there is a big difference in the other cells; therefore there should be a relationships (Strategy 5.2, Student 1).

Finally, among the level 5 correct strategies we found students who compared two conditional distributions by row; for example, *a/(a+b)* with *c/(c+d)* or else compared conditional distributions by columns: "*When we observe the table, 60% of people with insomnia have stress and also 60% of people with no insomnia have stress; the percentage is the same*" (Strategy 5.3, Student 28). Another correct strategy is comparing odds in favour and against *B* for each value of *A*; which was described by Batanero et al. (1996): "*There are 90 people with insomnia for every 60 with no insomnia when you have stress; that is the odds are 3/2; the same odds 60/40 apply when you do not have stress*" (Strategy 5.4, Student 21).

In Table 6 we present the frequency of responses in the above categorization and percent of students in the sample. As regards levels of complexity, students tended to use either level 2 or level 4 strategies none of which are correct, although part of them are partly correct and helped the students to get a correct association judgment. There was a big percentage of students who did not use all the cells information, since their strategies were level 3 or lower.

About 26% of students who used level 4 strategies compared joint frequencies among them, an incorrect strategy described by Batanero et al. (1996) and about 74% of them used the four cells with additive comparisons, a partly correct strategy described by Inhelder and Piaget (1955) in the concrete-operation level but that also was found by Batanero et al. (1996) in high school students. Finally most of level 5 strategies were correct as students either compared the odds ratios (relative risks) or compared conditional distributions a strategy proposed by Jenkins and Ward (1965) and also found in Batanero et al. (1996).

Level	Correctness	Item 1 (Independence)		Item 2 (Inverse)		Item 3 (Direct)	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
Level 0	Incorrect	13	3.1	15	3.6	15	3.9
Level 1	Incorrect	73	17.6	20	4.8	33	8.0
Level 2	Incorrect	108	26.1				
Level 2	P. correct			154	37.2	153	37.0
Level 3	Incorrect	27	6.5				
Level 3	P. correct			16	3.9	9	2.2
Level 4	Incorrect	27	6.5	12	2.9	19	4.6
	P. correct	76	18.4	115	27.8	100	24.2
Level 5	Incorrect	20	4.8	29	7.0	25	6.0
	P.correct	10	2.4	8	1.9	6	1.4
	Correct	46	11.1	37	8.9	28	6.8
Do not explain their strategy		14	3.4	8	1.9	25	6.0
Total		414		414		414	

Table 6. Frequency of strategies (and percent of students) by item, level and correction

To conclude, there was a scarce use of correct strategies (11%, 8.9 and 6.8 depending on the item). In spite of this, most students correctly perceived the association in items 2 and 3 and their estimation of association was reasonably accurate. The explanation is that an important percentage of students used partly correct level 2, 3 and 4 strategies in these two items, where these partly correct strategies are used to obtain a correct association judgement. Specifically, 79.7% of students used partly correct strategies in Item 2 and 71.6 % in Item 3, while in Item 3 the percentage was only 31.9%. These different percentages, added to the illusory correlation phenomena explaining the different performance in the three items.

The mean estimate of the association coefficient in according to the strategy correctness (incorrect, partly correct or correct strategy) is presented in Table 6 with the typical error. In the last two columns of this table we present the association coefficient (Phi value) in each item. We observe that the estimated value of the association coefficient is closer to zero when the strategy is more correct; in particular the estimated average value is very close to zero for correct strategies. This result suggest that partly correct and correct strategies help students perceive independence in the data in this item, even when their previous beliefs about the association were contrary to independence.

In the items 2 and 3, the average estimated value of the association coefficient increase with the strategy correctness and this fact suggest that partly correct and correct strategies helped the students perceive that these data presented high-moderate association. The differences in the estimated average values according the type of strategy (incorrect, partly correct or correct) were statistically significant in the Anova test of in all the three items (p values included in the last column in Table 7.

Item	Incorrect strategy		Partly correct strategy	Correct strategy			Association coefficient	p value (Anova)
	Mean	Typical error	Mean	Typical error	Mean	Typical error		
1	0.536	0.012	0.432	0.024	0.174	0.038	0	0.000
2	0.655	0.029	0.743	0.011	0.809	0.025	-0.62	0.001
3	0.665	0.026	0.716	0.013	0.756	0.029	0.67	0.035

Table 7. Mean estimate of association coefficient in different type of strategy

Semiotic Conflicts in Interpreting 2x2 Tables

The high percentage of students declaring that the variables are associated in Item 1 (78.1%; Table 3), consistently with a moderate-sized average estimated coefficient (.47; Table 4), shows that part of these students reasoned according to illusory correlation. Their incorrect perception of association may also be explained by the wide use of incorrect strategies (64.7) in this item.

To follow we use the idea of semiotic conflict to explain the high percentage of incorrect strategies in Item 1. Our suggestion is that part of the students using these incorrect strategies confused or misinterpreted some of the mathematical objects implicit in 2x2 tables; or else attributed to them some incorrect properties.

To check this conjecture we performed a further detailed qualitative analysis of the students' responses linked to incorrect strategies in Item 1. This analysis helped us to identify these latent semiotic conflicts. To follow we classify these conflicts, according to whether they involve incorrect properties assigned to association or incorrect properties assigned to independence. We include a typical response in each category

Incorrect properties assigned to association. In order to justify association in Item 1, some students assigned non-existent properties to association. The more frequent incorrect properties of association assumed by the students were the following:

•*Identifying association and causality:* Although causality always involves association, association does not always involve causality; but some students misinterpreted that this relation was symmetrical. This belief was also found in Batanero et al. (1996) who assumed it was a stable conception (causal conception of association). In our study this belief appeared in all the level 0 strategies, not only in Item 1, but also in items 2 and 3.

•*Assuming that association can be deduced from only a part of the data.* Association is a property of the data distribution and not a property of some isolated values of the distribution. However, in level 3 or lower level strategies, the students' discard some data in the table to judge the association. Using only a part of the data to solve 2x2 table association problems was also described by Smedlund (1963), Beyth -Marom (1982); Shaklee and Mins (1982), Yates and Curley (1986) and Pérez-Echeverría (1990). Depending on the table particular data Level 2 and 3 strategies may work for the particular problem, as happened in our study. For items 2 and 3 these strategies were partly correct and provided a successful response. However, in Item 1, students who analyzed only a part of the data usually led to a judgment of positive association, as well as to a moderate or high association coefficient.

•*Assuming that there is association when the absolute frequency in the distribution of a variable, when conditioned by the other variable changes.* We can see this belief in those students who deduced association from additive comparisons (Level 4 strategies). This strategy should have been overcome by our students, since according to Inhelder and Piaget (1955) is prior to the formal operations stage. However, it appeared in our sample in Strategy 4.2. Students using this procedure only took into account the favourable cases (and not all the possible cases) when comparing probabilities. Therefore this strategy involves *a conflict* in understanding the idea of probability or confusion between favourable cases and probability. Again the strategy worked well in items 2 and 3, but not in Item 1.

•*Assuming that a difference between the sums of diagonals in the 2x2 table involves association in the data.* This strategy was considered to be correct by Piaget and Inhelder, but Allan & Jenkins (1983) and Shaklee and Mins (1982) suggested it does not work in the general case as we see in Item 1, where Strategy 4.1 lead to an incorrect judgment. However it worked well for items 2 and 3 in our study.

•*Assuming that a>d in the case of association*. This incorrect property appears in Strategy 2.2, where students did not understand that cell *d,* has the same relevance as cell *a* on the association. We did not find this strategy in previous research.

Incorrect properties attributed to independence. In addition to attributing nonexistent properties to association, other students required incorrect properties to decide that there was perfect independence in the data. In particular we observe some students who expected identical joint frequencies in the case of independence (Strategy 5.2). This belief also involves some confusion between the ideas of independence and equiprobability and was not described in previous research.

Discussion and Implication for Teaching

Most psychology students in our study judged association, even in cases where there was none and our results were worse than those in Batanero et al. (1996) in perceiving independence. However our students had a better perception of negative association. These authors did not inform about the estimate of association by their students so that we provide new information about this point. In our study the estimation was very accurate in the case of association and poorer in the case of independence. The estimation was consistent with the association judgment. Results were very close in all participating universities, which suggest that students' reasoning and beliefs as regards association were similar in these different educational contexts.

The illusory correlation phenomenon and the students' previous beliefs may have influenced the students' association judgment and their accuracy in estimating the association coefficient in Item 1 (independence). However, the high percentage of incorrect strategies in this item suggested additional problems in the understanding of the ideas of association and independence. The students tried to justify their previous belief that there was association in this item, using a variety of incorrect strategies. The reasoning behind their explanations revealed a number of incorrect properties assigned to either association or independence that, on one hand served to justify their incorrect strategies and on the other hand confirmed their previous expectation in the association of the data.

We observed the causal conceptions described by Batanero, Estepa, Godino and Green (1996), but not the unidirectional conception, since most students in our study perceived the association when this was negative. In addition we listed some new semiotic conflicts related to misinterpreting mathematical objects and attributing incorrect properties to the ideas of association and independence.

For example, some students assumed that association may be judged from only part of the data, and did not perceive association as a property of the distribution. A high percentage of students tried to deduce the existence of association basing their conclusions only on absolute frequencies. This behaviour was also described by Konold, Pollatsek, Well y Gagnon (1997), who suggested that, when given two groups to compare students rarely used a statistically appropriate method of comparison. In particular, in our study many students tried to compare two groups in the 2x2 table using frequencies rather than percents to make the groups comparison.

Other students rejected association in the data when the frequencies in cells *b* or *c* was not null. They expected a determinist relationship between the variables; however an association between two variables indicates a tendency but we cannot be certain how the association will affect each individual. Even a moderate association allows doctors and psychologists to provide advice to people in danger of some risk; but unless the association is perfect they cannot make exact predictions about what will happen to individuals as a result of a given behaviour (Nunes, & Bryant, 2011).

Finally, we noticed that some students required that the four cells in the table should have equal frequency in case of independence; these students showed an intuitive idea of what is an expected frequency, but however, they confused independence and equiprobability.

According to Schield (2006), an educated person should be able to critically read tables in the press, Internet, media, and professional work. This involve not only the literal reading, but being able to identify trends and variability in the data, including the correct judgment of association. All these reasons and our results suggest the need for further research about teaching association, since the causal conception and the effect of illusory correlation does not seem to improve with traditional instruction (Batanero, Godino, & Estepa, 1998). Since semiotic conflicts do not assume a strong conviction on the part of the students, it is possible to change them with adequate instruction and then the identification of these conflicts in the students is a first step in order to correct their wrong reasoning and improve their competence in judging association.

According to Inhelder and Piaget (1958) and Adi, Karplus, Lawson, and Pulos (1978), understanding association depends on understanding probabilities and proportions. These two types of competencies should be developed in our students so that they can progress in correlational reasoning. Moreover, in order to draw inferences from the frequencies in 2x2 tables, students must understand the relevance of the different cells in Table 1 for a mutual relationship between the variables (Nunes, & Bryant, 2011); although students easily recognise the relevance of cell *a* for the association it is harder that they view the relevance of cell *d*. Nunes and Bryant (2011) suggest that understanding the relevance of call *d* for the association requires reasoning about a contradiction; this is not simple in deterministic situations and less in correlational situations. Moreover these authors also suggest that correlation reasoning requires understanding randomness and assessing the extent to which the association we perceive between the table frequencies departs from what could be expected by chance.

This formation is particularly needed in the statistics education of professionals like psychologist who not only should assess association in risk situations, but only communicate this information to their clients. For example Gigerenzer and Edwards (2003) indicate that even simple probabilities are a steady source of miscommunication because we often leave open the sample space or the population to which the probability refers. The same ambiguity occurs in communicating clinical risk, and the associate factors, such as the side effects of a drug. For this reason 2x2 tables, when properly understood and interpreted may turn in an useful tool for psychologists.

All these abilities should be developed in the students with a careful planning of teaching, Our purpose is to continue this work by designing teaching activities that confront students with their biases and help them overcome them as well as improve their partly correct strategies.

Acknowledgements

Research supported by the project EDU2013-41141-P (MEC, Spain) and group FQM126 (Junta de Andalucía).

References

Adi, H., Karplus, R., Lawson, A., & Pulos, S. (1978). Intellectual development beyond elementary school: correlational reasoning. *School Science and Mathematics, 80*(8), 675-683.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation, 14* (4)*,* 381-405.

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information, *Psychological Review, 91* (1), 112-149.

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112* (1), 117-135.

Barbancho, A. G. (1992). *Estadística elemental moderna* (Modern elementary statistics). 15th edition. Barcelona: Ariel.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. (1996) Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*(2), 151-169.

Batanero, C., Godino, J. D., & Estepa, A. (1998). Building the meaning of statistical association trough data analysis activities. In A. Olivier y K, Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (v.1, pp. 221-242). Stellembosch, South Africa: International Group for the Psychology of Mathematics Education.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and cognition*, *10* (6), 511-519.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6* (1), 151-155

Chapman, L. J. & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs, *Journal of Abnormal Psychology, 74* (3), 271-280.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90* (2), 272-292.

Díaz, J., & Gallego, B. (2006). Algunas medidas de utilidad en el diagnóstico (Some useful measures in diagnosis). *Revista Cubana de Medicina General Integrada, 22*(1). Retreived from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252006000100008.

Eco (1979). *Tratado de semiótica general* (General semiotics). Barcelona: Lumen.

Einhorn, H. J., & Hogart, R. M. (1986). Judging probable cause. *Psychological Bulletin*. *99* (1)*,* 3-19.

Erlick, D. E., & Mills, R.G. (1967). Perceptual quantification of conditional dependency, *Journal of Experimental Psychology, 73* (1)*,* 9-14.

Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores* (Preconceptions on association and its evolution with computer-based teaching). Unpublished Ph.D. University of Granada, Spain.

Estrada, A., & Díaz, C. (2006). Computing probabilities from two way tables. An exploratory study with future teachers. In A. Rossman, & B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics.* Salvador (Bahia): International Association for Statistical Education. Retrieved from: http://iase-web.org/Conference_Proceedings.php?p=ICOTS_7_2006.

Falk, R. (1986). Conditional probabilities: insights and difficulties. In R. Davidson, & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. (pp. 292-297). Victoria, Canada: International Statistical Institute. Retrieved from: http://iase-web.org/Conference_Proceedings.php?p=ICOTS_2_1986.

Font, J. D., Godino, J. D., & D'Amore, B. (2007). An ontosemiotic approach to representations in mathematics education. *For the Learning of Mathematics, 27* (2), 3-9.

Gigerenzer, G. (2003). *Reckoning with risk: learning to live with uncertainty.* London: Penguin.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal, 327*(7417), 741-744.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological science in the public interest, 8*(2), 53-96.

Inhelder, B., & Piaget, J. (1955). *De la logique de l´enfant à la logique de l´adolescent.* (From the child's logic to the adolescent's logic). Paris: Presses Universitaires de France.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of the contingency between responses and outcomes, *Psychological Monographs, 79* (1)*,* 1-17.

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). New York: Cambridge University Press.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997) Students analyzing data: Research of critical barriers. In J. Garfield, & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 151-16). Voorburg, The Netherlands: International Statistical Institute.

Martignon, L. (2014). Fostering children's probabilistic reasoning and first elements of risk evaluation In E. J. Chernoff, B., & Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 149-160). Dordrecht: Springer.

Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology, 363*(3), 315-336.

Nunes, T., & Bryant, P. (2011). Understanding risk and uncertainty: the importance of correlations. *EM TEIA, 2*(2). Retrieved from: www.gente.eti.br/revistas/index.php/emteia/.

Pérez-Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico* (Psychology of probabilistic reasoning). Madrid: Ediciones de la Universidad Autónoma.

Power, M. (2007). *Organized uncertainty: Designing a world of risk management.* Oxford; Oxford University Press.

Schield, M. (2006). Statistical literay survey analysis: reading graphs and tables of rates percentages. In A. Rossman, & B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics.* Salvador (Bahia): International Association for Statistical Education. Retrieved from: http://iase-web.org/Conference_Proceedings.php?p=ICOTS_7_2006.

Shaklee, H. (1983). Human covariation judgment: accuracy and strategy. *Learning and Motivation, 14* (4), 433-448.

Shaklee, H., & Mins, M. (1982). Sources of error in judging event covariations: Effects of memory demands, *Journal of Experimental Psychology Learning, Memory and Cognition, 8*(3), 208-224.

Slovic, P. E. (2000). *The perception of risk.* London: Earthscan Publications.

Smedlund, J. (1963). The concept of correlation in adults. *Scadinavian Journal of Psychology, 4* (1), 165-174.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.

Wright, J. C, & Murphy, G. L. (1984). The utility of theories in intuitive statistics: the robustness of theory-based judgments, *Journal of Experimental Psychology General, 113*(2), 301-322.

Yates, J. F. & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement, *Acta Psicológica, 62* (3), 293-302.

Zieffler, A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course.* Unpublished PhD. University of Minnesota.

Appendix: Questionnaire

Item 1. A researcher is studying the relationship between stress and insomnia. In a sample of 250 people he observed the following results:

	Stress disorders	No stress disorders
Insomnia	90	60
No insomnia	60	40

a. Looking to these data, do you think there is a relationship between stress and insomnia? Explain your response.
b. Please mark on the scale below a point between 0 (minimum strength) and 1 (maximum strength), according the strength of relationship you perceive in these data.

0 0.5 1

Item 2. A psychologist got the following data to study the possible association between being an only child and being a problematic child:

	Problematic child	No problematic child
Having some brothers /sisters	40	100
No brothers or sisters (only child)	100	10

a. Looking to these data, do you think there is a relationship between being an only child and being a problematic child? Explain your response
b. Please mark on the scale below a point between 0 (minimum strength) and 1 (maximum strength), according the strength of relationship you perceive in these data.

0 0.5 1

Item 3. In order to assess possible association between sedentary life (not performing physical exercise) and allergy the following data were obtained:

	Suffering allergy	Do not suffer allergy
Sedentary life	130	30
No (Active life)	20	120

0 0.5 1

a. Looking to these data, do you think there is a relationship between sedentary life and suffering allergy? Explain your response.
b. Please mark on the scale below a point between 0 (minimum strength) and 1 (maximum strength), according the strength of relationship you perceive in these data.