*Eurasia Journal of Mathematics, Science & Technology Education,* 2015, **11**(4), 735-750

# Comparing Psychology Undergraduates' Performance in Probabilistic Reasoning under Verbal-Numerical and Graphical-Pictorial Problem Presentation Format: What is the Role of Individual and Contextual Dimensions?

Mirian Agus
*University of Cagliari, ITALY*
Maribel Peró-Cebollero
*University of Barcelona, SPAIN*
Maria Pietronilla Penna
*University of Cagliari, ITALY*
Joan Guàrdia-Olmos
*University of Barcelona, SPAIN*

*Received 30 December 2014; accepted 6 May 2015*

This study aims to investigate about the existence of a graphical facilitation effect on probabilistic reasoning. Measures of undergraduates' performances on problems presented in both verbal-numerical and graphical-pictorial formats have been related to visuo-spatial and numerical prerequisites, to statistical anxiety, to attitudes towards statistics and to the confidence in response correctness. Psychology undergraduates in Italy and Spain with no statistical expertise (N= 676) completed a protocol under conditions of presence versus absence of time pressure. Hierarchical linear regressions and ANCOVAs with mixed design have been carried out separately for each sample. The best predictor of performance in both formats has been the confidence in solution correctness under the condition of time pressure administration, which seemed to promote the commitment to the task. The findings suggest that the eventual occurrence of a graphical facilitation could be the result of a multifactorial interaction among contextual and individual dimensions, rather than being strictly related to the problem presentation format.

*Keywords*: confidence in performance correctness, numerical ability, probabilistic reasoning, problem presentation format, statistical anxiety, visuo-spatial ability

## INTRODUCTION

Many studies have sought to appraise the effects on problem solving performance of the individual and task features. In particular, probabilistic problems belong to a specific kind of task which subjects must cope with on a daily basis, chiefly when dealing with statistics. As it is well known, probabilistic problem solving implies multifaceted relationships among many elements, including previous experiences and performances (Galli, Chiesi, & Primi, 2011; Guàrdia-Olmos et al., 2006), beliefs and attitudes towards statistics (Garfield & Gal, 1999), emotional aspects (Chiesi, Primi, & Carmona,

Correspondence to: Mirian Agus,
Department of Pedagogy, Psychology, Philosophy, Faculty of Humanistic Studies. University of Cagliari, Italy.
E-mail: mirian.agus@unica.it
doi: 10.12973/eurasia.2015.1382a


ISSN: 1305-8223

### State of the literature

- In the last decades the real difficulties faced by undergraduates in statistics have been highlighted. Specifically, the study of probabilistic reasoning evaluated the characteristics of this reasoning, searching for ways to improve the performances of participants in these problems.
- The presentation of problems in frequency formats, in a natural sampling framework, using graphical-pictorial representations has been evaluated as a useful way to enhance performance.
- Nevertheless, it is necessary to clarify how it is possible to explain these enhancements. The relations among multiple contextual and individual factors deserve further investigation, specifically in relation to the contribution of cognitive and non-cognitive dimensions.

### Contribution of this paper to the literature

- The paper clarifies the features of graphical facilitation on probabilistic reasoning, their relations with contextual and individual characteristics, in Italian and Spanish undergraduates without any statistical knowledge.
- We compare the performance in verbal-numerical and graphical-pictorial formats, in presence versus absence of time pressure, regarding visuo-spatial and numerical abilities, anxiety, attitudes and confidence.
- Differently from bibliography, we matched the performance of the same student in similar problems administered in both presentation formats. There are not researches that considered simultaneously the effects of all these dimensions; specifically the role of visuo-spatial abilities, confidence and time pressure have largely been disregarded.

2011) and metacognitive processes (Morony, Kleitman, Lee, & Stankov, 2013).

Concerning probabilistic reasoning, based on processes which are essential in understanding and using statistics (Ben-Zvi & Garfield, 2004; Evans, Thompson, & Over, 2015; Mandel, 2014a), many studies tried to evidence the so-called effect of graphical facilitation. In this regard a number of authors made different, often contrasting, claims (e.g., Brase & Hill, 2015; Moro & Bodanza, 2010).

In absence of a detailed general theory of statistical reasoning, we can resort only to generic considerations, like the ones stemming from the dual-coding theory (Paivio, 1971), implying that the graphical-pictorial representations could better support probabilistic reasoning. Indeed, the solution of probabilistic problems should involve the construction of a mental model, which, in turn, would be related not only to verbal but also to visuo-spatial processing. Johnson-Laird et al. (Johnson-Laird, 1983; Johnson-Laird, Khemlani, & Goodwin, 2015) extended the area of theory's application, stating that subjects solve problems by applying higher-level representations. Within this context, one consequence could be that the observed failures should be related to the improper selection of the specific mental model. Consequently, the use of problem presentation formats supporting the correct selection of the needed mental model would help to improve the individual's performance.

In order to assess the validity of these generic ideas in specific contexts, recent studies have investigated the use of graphical representations in many areas of knowledge. The aim has been the understanding of the relationships among different reasoning dimensions (Britt & Chen, 2013; Cumming, 2007; Fernández-Aguirre, Garín-Martín, & Modroño-Herrán, 2014). Some authors asserted that graphic formats may provide a way to clarify probabilistic reasoning (e.g., Brase, 2000; 2009; Brase & Hill, 2015; Johnson & Tubau, 2013; Mandel, 2014b; Tubau, 2008). Conversely, other authors (delMas, Garfield, & Ooms, 2005; Knauff & Johnson-Laird, 2002) have argued that graphic formats may obstruct reasoning because the cognitive system could thereby be loaded with auxiliary indications that are irrelevant to problem solving.

In this regard it could be suitable to mention the Stanovich's problem solving Dual-Process Theory (2009; 2011). The latter supposes that problem data are managed by two kinds of processing: autonomous (Type I) and analytic (Type II). The Type I is connected to activities defined as non-conscious, contextualised, associative and automatic, which don't require an intense use of the working memory. Conversely, the Type II processing is characterized by a robust application of working memory, the normative responses, precise and rule-based, in turn associated with general and specific abilities, as well with plain knowledge (Evans & Stanovich, 2013; Klaczynski, 2014). The Type II designates the algorithmic mind (demarcated by individual differences in fluid intelligence, linked to the education) and the reflective mind (defined by metacognitive procedures that establish the rational thinking dispositions) (Evans & Stanovich, 2013). In principle, this perspective could yield a useful key to explore the eventual occurrence of facilitation effects. Namely it could take into account the interactions between different aspects of problem solving and the possibility that dissimilar formats of problem presentation could support different types of data processing processes (e.g., Klaczynski, 2014; Oaksford, 2015). However, in order to achieve a higher predictive power within the specific context of facilitation effects, the Dual Process Theory must be complemented by more particular hypotheses about the mental coding processes which could be more important when searching for the solution of a probabilistic problem.

In this regard we find in literature two different positions. The first one is given by the natural frequency hypothesis (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Zhu & Gigerenzer, 2006) asserting that subjects search for a solution by building a mental model of the relationships among the different data frequencies occurring in the problem at hand. Therefore the graphical devices could help the subject to building the more effective model of frequencies relationships. Instead the *nested-set* hypothesis (e.g. Evans, Handley, Neilens, & Over, 2010; Girotto & Gonzalez 2001; Sloman, Over, Slovak, & Stibel, 2003; Yamagishi, 2003) holds that the mental model to be built by a subject is made by the relationships among the different sets and subsets of data occurring within the problem. In this case the graphical representation might support the subject to detect the more effective structure of subset relationships.

Recently Sirota et al. (Juanchich & Sirota, 2015; Sirota & Juanchich, 2011; Sirota, Juanchich, & Hagmayer, 2014) evaluated the contribution of individual differences in Bayesian reasoning, sustaining the *nested-set* approach in the clarification of this reasoning. These authors moreover corroborated the null iconicity effect in order to explain the facilitation effect in probabilistic reasoning (Sirota, Kostovičová, & Juanchich, 2014), related even now to the *nested sets* approach. Formerly the facilitation effect in the probabilistic reasoning might be related to the structure of the problem presentation and to the use of frequency format, rather than to the use of graphical devices (e.g., Sirota, Kostovičová, & Juanchich, 2014). For these reasons in the study of performance in probabilistic reasoning it appears as more convenient to investigate the individual differences (Sirota, Juanchich, & Hagmayer, 2014; Stanovich, & West, 2000) and the contribution of specific features, subject's abilities and beliefs (Hoffman, 2010).

In reference to these perspectives, it would be challenging to identify the specific individual and task features that could support the effect of graphical facilitation. The experimental evidence thus far collected in probabilistic problem solving is conflicting. A further investigation involving other relevant variables, in addition to the presentation mode, is therefore necessary to assess the part played by graphics in modulating the performance. This paper reports some preliminary results of one such study.

## The Effect of Metacognitive, Cognitive, and Non-Cognitive Factors on Probabilistic Reasoning

Many studies, attempting to investigate the multiple dimensions involved in this reasoning, have been conducted.

For instance, Lalonde and Gardner (1993) identified some predictors (as well as their interactions) of outcomes in statistical achievement: mathematical abilities, attitudes, statistical anxiety and effort. Tremblay, Gardner and Heipel (2000) showed that anxiety, aptitudes and motivational aspects can affect statistical achievement. Additionally, Sorge and Schau (2002) evidenced the influence of previous curricula and attitudes towards statistics on statistics achievement. Chiesi and Primi (2009a; 2010) observed that mathematical ability was a crucial factor in attitudes towards and performance in statistics in Psychology undergraduates.

Visuo-spatial ability appears as crucial when evaluating the effects of graphical facilitation, but few studies have been conducted on this topic. Kellen, Chan and Fang (2006; 2007; 2013) evaluated the impact of diagrams on reasoning, and observed that the presence of graphics increases the difficulty in problem solving for individuals with low visuo-spatial abilities. They supposed that these subjects may have difficulties when solving problems in graphical-pictorial format because the cognitive process could share working memory resources with non-visual reasoning, thus overloading the capacity for information processing (Kellen et al., 2013). They also emphasized that performance in the verbal-numerical format was not affected by the level of visuo-spatial ability, suggesting that problems should be presented in simple verbal-numerical format for subjects with low visuo-spatial abilities, and in simple graphical-pictorial format for individuals with high visuo-spatial abilities (Kellen et al., 2013).

Another relevant variable to be considered is confidence. Some authors (Harvey, 1997; Jackson & Kleitman, 2013; Moore & Healy, 2008) have highlighted that post item judgement confidence may influence the subject's metacognitive processes in problem solving. Other authors (Morony et al., 2013; Stankov, 2013; Stankov, Lee, Luo, & Hogan, 2012) showed that confidence is the best predictor of achievement in mathematics with respect to self-efficacy and anxiety. This aspect has been studied also by other authors (e.g., Thompson, Prowse, & Pennycook, 2011; Evans, Thompson, & Over, 2015) that identified the role of feeling of rightness in the problem solving and its relationship with the Type I and Type II data processing in the dual-processes approach (Stanovich, 2009).

Then, referring to the above quoted papers, the interaction between the previously listed dimensions appears to be relevant when trying to account for graphical facilitation effect. More specifically, the investigation about this interaction seems to be more promising when dealing with students in the humanities who are naïve in statistics. Among these students, the difficulties faced by Psychology undergraduates in statistics are renowned. Guàrdia-Olmos et al. (2006) affirmed that essentially, the problem is tied to the excessive differences in academic backgrounds and the unrealistic expectations of the first-year undergraduates, who do not expect to encounter mathematics in their coursework. Indeed, they often feel inadequately prepared to cope with this content, or do not meet the requirements necessary to approach the matter (Chiesi & Primi, 2009a; Guàrdia-Olmos et al., 2006). All of these aspects could have a negative impact on students' performance in statistics, which could become a significant obstacle for their studies.

In summary, this paper aims to explore the role of visuo-spatial and numerical abilities, attitudes towards statistics, statistical anxiety and confidence on students' performance in probabilistic reasoning, assessed in correspondence to verbal-numerical (N) and graphical-pictorial formats (G). These evaluations may shed light on the effect of graphical facilitation identified in the bibliography (Brase & Hill, 2015; Kellen et al., 2013; Garcia-Retamero, Galesic, & Gigerenzer, 2011; Moro, Bodanza, & Freidin, 2011).

In this compound frame, time administration in problem solving appeared to be a crucial aspect. The timing could have an important effect on performance in connection to the cognitive and non-cognitive factors under consideration (Beilock, Kulp, Holt, & Carr, 2004). The presence of time limits could support students' commitment to or increase their misuse of solution strategies (Beilock & DeCaro, 2007; Maule, Hockey, & Bdzola, 2000; Rieskamp & Hoffrage, 2008). For this reason, we assessed student performances in both formats under conditions of presence versus absence of time pressure.

Moreover we attempted to inquire into these relationships in two Mediterranean countries (Italy and Spain), both of which are included in the European Higher Education Area (EHEA). These countries are frequently subject to comparison, showing both their differences and their similarities (e.g., Agasisti & Cordero-Ferrera, 2013). We seek to explore whether the relationships among the cited variables are affected by the specificities of the two educational systems.

The variables assessed in this work are related both to the structure of the task and to individual idiosyncrasies. Thus, the innovation of this investigation concerns the merged evaluation of all these aspects, which have frequently been observed disjointedly in other studies.

## METHOD

### Participants

Six hundred and seventy six first-year undergraduates in Psychology aged 17 to 52 years (M= 20.32 years, SD = 6.09), attending university in Italy (N= 549, males 27.1%) and Spain (N= 127, males 28.3%) participated in the research. They are enrolled in the Italian universities of Cagliari (n= 28), Chieti (n=144), Genoa (n=58), Milan (Catholic University n=50), Naples (Second University n=35), Pavia (n=20), Rome (Sapienza University n=173) and Trieste (n= 41); in Spain, the study was administered at the University of Barcelona. In the Italian sample, the Rome undergraduates completed the protocol under the condition of time pressure, unlike all other students. All subjects voluntarily participated in the research and gave their consensus to using their data for an empirical study.

The administration of the protocol in the different cities was based on the availability of each class. The research was conducted applying non-probabilistic sampling (convenience sampling). Data collection was completed in the first semester, from September 2013 to January 2014.

All research participants were undergraduates without statistical skills who had never previously studied statistics in their curricula.

### Instruments

The undergraduates' demographic features were assessed, exploring gender, age, and previous curriculum exposure. Subsequently, each participant was presented with different sections, inquiring into visuo-spatial and numerical abilities (A and B), statistical reasoning in verbal-numerical and graphical-pictorial formats (C and E), attitudes towards statistics (D) and statistical anxiety (F).

In sections A and B, the Intermediate Form of Primary Mental Abilities (PMA) (Thurstone L.L. & Thurstone T.G., 1981; 1987) was applied, using the visuo-spatial and numerical scales. This instrument has been widely used with undergraduates (e.g., Fischer, Hickey, Pellegrino, & Law, 1994) because it is an effective instrument to assess spatial relations (Colom, Contreras, Botella, & Santacreu, 2002; Hegarty & Kozhevnikov, 1999), or the ability to perceive spatial patterns (useful in understanding graphical representations).

The C and E sections consisted of items inquiring into probabilistic reasoning in two formats (N and G), appointed in previous phases of research (Agus, Peró-Cebollero, Guàrdia-Olmos, & Penna, 2014; Agus, Peró-Cebollero, Penna, & Guàrdia-Olmos, 2015). These sections included five problems in both formats that questioned simple and conditional probabilities, referring to fundamental mathematical knowledge learned in high school. The G format included tree diagrams and iconic drawings, referring to classical studies conducted on these topics (e.g., Cosmides & Tooby, 1996; Sloman, Over, Slovak, & Stibel, 2003; Moro & Bodanza, 2010; Yamagishi, 2003). All problems avoided the use of any irrelevant information that was not closely related to the structure of the problem (e.g., Knauff & Johnson-Laird, 2002; Moro et al., 2011). Two examples of problems are shown in Figure 1 (N format) and Figure 2 (G format).

Each item comprised a brief description and four closed response options (with one correct response). An open-ended question was also included whereby the undergraduates had to explain the reasoning they applied to solve the problem. A question inquiring into the students' confidence about the correctness of their performance (Likert scale: 1 = "no confidence", to 5 = "totally confident") was associated with each problem (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Morony et al., 2013; Sloman et al., 2003). The psychometric characteristics of the instrument were previously evaluated (Agus et al., 2015), and the Italian and Spanish versions were created by means of back translation. The performances were assessed by adding the total number of correct responses in both formats (Agus et al., 2015).

Section D assessed the Survey of Attitudes Towards Statistics (SATS-28 - Schau, Stevens, Dauphine, & Del Vecchio, 1995). Students were asked to indicate their responses (7-point Likert scale ranging from 1 = "strongly disagree", to 7 = "strongly agree"). The SATS implies four dimensions: Affect (six items); Difficulty (seven items); Cognitive Competence (six items); and Value (nine items). The Italian version (Chiesi & Primi, 2009b) and Spanish version (Carmona-Màrquez, 2004) were administered in a way consistent with the original English questionnaire, and cross-country validity in Italy and Spain was demonstrated. The SATS Cronbach's α

A factory produces electronic games, but not all of them work well. Of every 100 game products:

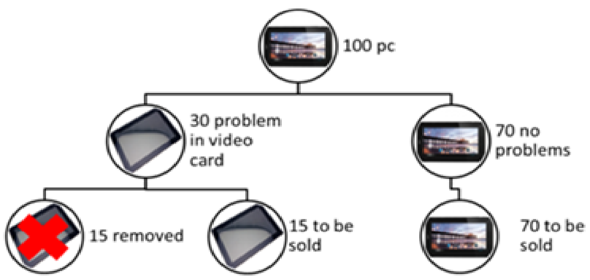
- 20 may have an electrical problem,
- 80 can work correctly.

The company has developed control systems to identify faulty games; however, these systems do not work properly. In reality, half of the games with electrical problems continue in the production line, where they are considered as well functioning.
If you randomly extract a game that has been sent to shops for commercialization and evaluated as free of defects, what is the probability that it is defective?

a) 10 / 90
b) 10 / 100
c) 10 / 80
d) 20 / 100

**Figure 1.** Item example – Verbal-numerical format

A factory that produces personal computers has problems in the production process. Some of the computers are defective (problems with the video card). Such problems are not always identified by the quality control and consequently some defective computers are sent forward in the production line. The graphic below shows this process. What is the probability that a computer sent to shops for commercialisation and evaluated as free of defects, is defective?




a) 15 / 100
b) 15 / 70
c) 15 / 85
d) 3 / 10

**Figure 2.** Item example – Graphical-pictorial format

were good for the Italian (from .60 to .81) and the Spanish version (from .64 to .90). The scores were calculated by adding the answers for all items on the basis of manual indication.

In section F, the Statistical Anxiety Scale (SAS - Colet, Seva, & Condon, 2008) was administered. This scale consisted of 24 self-report items that required students to describe how they feel, for example, when attending a statistics class or an exam on a five-point scale. SAS involves three 8-item dimensions: Examination, Asking for Help and Interpretation. The score computation was based on the sum of answers. The students were presented with the Italian (Chiesi et al., 2011) or the Spanish versions (Colet et al., 2008) based on their country. The measures showed high internal consistency (from .85 to .90 in the Italian version and from .81 to .92 in the Spanish version).

All instruments were administered in paper-and-pencil format to large groups in a quiet lecture room.

### Procedure

Each student was assessed on reasoning about problems in both N and G format in the same session in order to control the effect of individual features in the performance. All problems were administered randomly in different orders (NG, first N then G format, versus GN, first N then G format) and reversed sequences (1 and 2, the second reversed in respect to the previous, ensuring that each item was not in the same position on the protocol). Then, four modalities of problem presentation were obtained (24.4% of total sample answered the problems in the GN1 modality, 23.4% in GN2, 25.9% in NG1 and 26.3% in NG2).

When evaluating the answers, even cases in which the subject did not solve all problems were considered. Indeed, the missing answers often evidenced a great difficulty, as confirmed by the open-ended question following each problem.

Five hundred and three undergraduates participated in the protocol with no time limit (Italian n= 376 and Spanish n= 127) and 173 Italian undergraduates were subject to time limits. The latter were asked to complete sections C through F of the protocol within 30 minutes (for the PMA sections, the timing was fixed by the manual).

### Data Analysis

The data analysis was undertaken by using R (version 3.0.2) and SPSS (release 19) software.

First, the analysis was characterized by the description of all variables. Before verifying the research hypotheses, the potential order and sequence effects of problem presentation on performance were assessed. One Way Anova was carried out on the total number of correct responses in two formats (N and G). The variable identifying the order and sequence administration (NG1, NG2, GN1, and GN2) was the factor used to the ANOVA of each sample. In all cases, it did not ascertain a significant effect of order and sequence in two formats [Spanish without time pressure: N format $F=0.728,\ df= 3;123,\ p=.537$; G format $K\text{-}W = 5.517,\ df= 3,\ p=.138$ - Italians without time pressure: N format $F= 0.012,\ df= 3;\ 372,\ p=.998$; G format $F= 0.950,\ df= 3;\ 372,\ p=.417$ - Italians with time pressure: N format $F = 1.329,\ df= 3;\ 169,\ p=.267$; G format $K\text{-}W = 6.745,\ df= 3,\ p=.080$]. Then, the information related to different orders and sequences was investigated as if dealing with a sole sample.

Formerly, parametric statistical analyses were applied to verify the hypotheses related to graphical facilitation. To assess the specific role of all variables in predicting performance (in the N and G formats, respectively), two Multivariate Linear Regressions were performed; the accuracy of performance was used as criterion and the identified variables were inserted as predictors.

Then, a mixed design Ancova was computed, whereby the repeated measures were the performances in N and G, and the covariates were the significant predictors identified in the previous regressions. This analysis was suitable to compare the number of correct responses in the two formats, controlling the variables influencing the performances.

## RESULTS

The descriptive statistics in each sample were then examined (Table 1). The samples presented similar values in terms of socio-demographics variables; moreover, the number of women exceeded the number of men in all samples. In relation to the university admissions mark, we observed similar values in two Italian samples (Table 1). The numerical scale was similar in three samples, differently as in the visuo-spatial scale, SATS and SAS dimensions. Furthermore, regarding the probabilistic problems, a higher number of correct responses was observed in Italians who worked under time pressure. Conversely, the Italians who worked without time pressure showed (always at a descriptive level) lower confidence than in other two samples (see Table 1).

To explore the effect of the variables in probabilistic reasoning, a Multivariate Linear Regression (Backward method) was applied, using the number of correct responses in the N format (Table 2) as the criterion,

**Table 1.** Descriptive Statistics for Each Sample

| | SPANISH SAMPLE | ITALIAN SAMPLES | |
|---|---|---|---|
| | FULFILLED PROTOCOL WITHOUT TIME PRESSURE (S_NTP) | FULFILLED PROTOCOL WITHOUT TIME PRESSURE (IT_NTP) | FULFILLED PROTOCOL WITH TIME PRESSURE (IT_TP) |
| N | 127 | 376 | 173 |
| PERCENTAGE OF WOMEN | 71.7 | 78.7 | 60.1 |
| AGE MEAN | 20.320 | 20.140 | 19.690 |
| AGE STANDARD DEVIATION | 6.098 | 4.297 | 1.638 |
| LOWEST AGE | 17 | 18 | 18 |
| HIGHEST AGE | 52 | 62 | 36 |
| PERCENTAGE BY UNIVERSITY AFFILIATION | BARCELONA 100% | CAGLIARI 7.4%<br>CHIETI 38.3%<br>GENOA 15.4%<br>MILAN 13.3%<br>NAPLES 9.3%<br>PAVIA 5.3%<br>TRIESTE 10.9% | ROME 100% |
| UNIVERSITY ACCESS MARK MEAN [1] | 8.822 | 78.967 | 78.351 |
| UNIVERSITY ACCESS MARK STANDARD DEVIATION [1] | 1.142 | 10.551 | 10.981 |
| LOWEST UNIVERSITY ACCESS MARK | 6.333 | 60 | 60 |
| HIGHEST UNIVERSITY ACCESS MARK | 13.000 | 100 | 100 |

| | SPANISH SAMPLE FULFILLED PROTOCOL WITHOUT TIME PRESSURE (S_NTP) | | ITALIAN SAMPLE FULFILLED PROTOCOL WITHOUT TIME PRESSURE (IT_NTP) | | ITALIAN SAMPLE FULFILLED PROTOCOL WITH TIME PRESSURE (IT_TP) | |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD |
| PMA VISUO-SPATIAL SCALE | 23.937 | 11.819 | 19.726 | 10.460 | 23.595 | 11.695 |
| PMA NUMERICAL SCALE | 17.110 | 6.312 | 17.205 | 6.696 | 18.5000 | 6.173 |
| SATS AFFECT | 22.000 | 4.829 | 19.104 | 5.075 | 19.432 | 5.336 |
| SATS COMPETENCE | 28.354 | 4.914 | 24.848 | 5.218 | 25.845 | 5.894 |
| SATS VALUE | 47.118 | 9.298 | 43.181 | 8.761 | 44.214 | 9.167 |
| SATS DIFFICULTY | 27.110 | 5.770 | 24.688 | 5.910 | 25.880 | 6.164 |
| SAS EXAMINATION | 31.110 | 6.136 | 33.861 | 5.999 | 28.542 | 9.714 |
| SAS INTERPRETATION | 17.204 | 5.151 | 19.104 | 6.548 | 15.505 | 6.572 |
| SAS HELP | 16.141 | 7.220 | 18.511 | 6.972 | 16.226 | 8.074 |
| FN CORRECT RESPONSES | 2.000 | 1.480 | 1.742 | 1.442 | 2.277 | 1.472 |
| FG CORRECT RESPONSES | 2.622 | 1.345 | 2.535 | 1.377 | 2.572 | 1.435 |
| FN CONFIDENCE | 3.434 | 0.852 | 2.846 | 0.951 | 3.248 | 0.962 |
| FG CONFIDENCE | 3.736 | 0.818 | 3.023 | 1.002 | 3.530 | 0.893 |
| FN SOLVING TIME (h:m) | 0:12 | 0:03 | 0:14 | 0:07 | 0:12 | 0:05 |
| FG SOLVING TIME (h:m) | 0:12 | 0:06 | 0:13 | 0:06 | 0:12 | 0:05 |

*Note 1: University access is structured differently in Spain and Italy. In Italy, the diploma mark ranges from 60 to a maximum of 100; in Spain, the mark ranges from 0 to 14.*

**Table 2.** Linear Regression (backward method). Criterion variable: number of correct responses in N format

| | Block | $R^2$ (SE) | Variables | Beta | t | p | Tolerance | VIF |
|---|---|---|---|---|---|---|---|---|
| Spanish sample without time pressure (S_NTP) n=127 | 1 | .116 (1.391) | PMA VISUO-SPATIAL | .104 | 1.179 | .241 | .908 | 1.102 |
| | | | PMA NUMERICAL | .083 | .965 | .336 | .938 | 1.067 |
| | | | SAS_EXAMINATION | -.069 | -.685 | .495 | .694 | 1.441 |
| | | | SAS_INTERPRETATION | -.059 | -.552 | .582 | .614 | 1.630 |
| | | | SAS_HELP | .077 | .796 | .428 | .749 | 1.335 |
| | | | SATS_AFFECT | -.082 | -.783 | .435 | .633 | 1.581 |
| | | | SATS_COMPETENCE | -.031 | -.271 | .787 | .547 | 1.829 |
| | | | SATS_VALUE | .083 | .869 | .387 | .763 | 1.311 |
| | | | SATS_DIFFICULTY | .177 | 1.719 | .088 | .662 | 1.511 |
| | | | FN CONFIDENCE | .315 | 3.282 | .001** | .760 | 1.316 |
| | Last 9 | .133 (1.378) | FN CONFIDENCE | .326 | 3.874 | .0001** | .969 | 1.032 |
| Italian sample without time pressure (IT_NTP) n=376 | 1 | .228 (1.266) | PMA VISUO-SPATIAL | .053 | 1.074 | .284 | .887 | 1.128 |
| | | | PMA NUMERICAL | .053 | 1.079 | .282 | .899 | 1.112 |
| | | | SAS_EXAMINATION | .005 | .100 | .920 | .740 | 1.351 |
| | | | SAS_INTERPRETATION | -.210 | -3.155 | .002** | .493 | 2.029 |
| | | | SAS_HELP | .032 | .536 | .593 | .614 | 1.629 |
| | | | SATS_AFFECT | .028 | .492 | .623 | .656 | 1.525 |
| | | | SATS_COMPETENCE | .032 | .578 | .564 | .718 | 1.393 |
| | | | SATS_VALUE | .075 | 1.431 | .153 | .788 | 1.269 |
| | | | SATS_DIFFICULTY | -.057 | -.988 | .324 | .649 | 1.540 |
| | | | FN CONFIDENCE | .321 | 5.671 | .0001** | .677 | 1.477 |
| | Last 8 | .233 (1.261) | SAS_INTERPRETATION | -.188 | -3.609 | .0001** | .797 | 1.255 |
| | | | FN CONFIDENCE | .343 | 6.634 | .0001** | .808 | 1.238 |
| Italian sample with time pressure (IT_TP) n=173 | 1 | .343 (1.211) | PMA VISUO-SPATIAL | .048 | .556 | .580 | .840 | 1.191 |
| | | | PMA NUMERICAL | .160 | 1.828 | .071 | .834 | 1.199 |
| | | | SAS_EXAMINATION | -.176 | -1.584 | .116 | .517 | 1.934 |
| | | | SAS_INTERPRETATION | .065 | .572 | .569 | .497 | 2.012 |
| | | | SAS_HELP | .022 | .214 | .831 | .616 | 1.622 |
| | | | SATS_AFFECT | .152 | 1.576 | .118 | .685 | 1.460 |
| | | | SATS_COMPETENCE | .033 | .338 | .736 | .668 | 1.496 |
| | | | SATS_VALUE | .049 | .571 | .569 | .868 | 1.151 |
| | | | SATS_DIFFICULTY | -.059 | -.672 | .503 | .840 | 1.191 |
| | | | FN CONFIDENCE | .425 | 4.225 | .0001** | .631 | 1.586 |
| | Last 9 | .357 (1.198) | PMA NUMERICAL | .191 | 2.315 | .023* | .919 | 1.089 |
| | | | FN CONFIDENCE | .525 | 6.372 | .0001** | .919 | 1.089 |

*Note:* * p<.05; ** p<.001*

followed by the number of correct responses in the G format (Table 3). The predictors were students' abilities, attitudes, anxiety and confidence.

For Spanish undergraduates on the first step in the N format, the significant positive effect of confidence was observed; this significance was confirmed by the last step of regression (β=.326). In this final step, the value of $R^2$ appeared low (.133), showing that there was a variability that remained unexplained by the predictors. For Italian undergraduates working without time pressure in the N format, confidence (β=.343) and anxiety in interpretation (β= - .188) had a significant effect ($R^2$ =.233). Finally, for Italians working under time pressure in the N format, the significant effect of numerical ability (β= .191) and confidence (β= .525) were detected ($R^2$ = .357) (Table 2).

Evaluating the regressions carried out using the responses in the G format as the criterion, we highlighted similar trends to those shown in the N format. In the Spanish sample, there are three significant predictors in the last step: the SATS value (β=.222), the SATS difficulty (β=.189), and confidence (β=.410) ($R^2$ =.243). For the Italian sample working without time pressure in the G format, the significant predictors were the PMA visuo-spatial (β=.140), the SAS interpretation (β=-.206), and confidence (β=.338). In this case, the final $R^2$ was .228. The Italian sample working with time pressure showed a similar $R^2$ (.218), but only confidence (β=.475) was a significant predictor (Table 3).

For all samples in both problem formats, the significant predictor was consistently the confidence. Moreover, it was observed that the variance explained in the G format was low and very similar in the three samples ($R^2$ varying from .218 to .243), unlike the variance in the N format, which had a larger range across the samples ($R^2$ from .133 to .357).

Mixed design ANCOVAs were then carried out to evaluate graphical facilitation, controlling the effect of significant predictors in the former regressions (inserted in the ANCOVAs as covariates). These variables were different in each sample examined; different analyses were thus applied. In particular, for the Spanish undergraduates, the covariates were the SATS Value, difficulty, and the confidence in N and G formats. For the Italian sample working without time pressure, the covariates were the SAS Interpretation, PMA visuo-spatial abilities, and confidence in both formats. For the Italian sample working under time pressure, the covariates were the PMA numerical abilities and confidence in the N and G formats (Table 4).

In the Spanish sample an effect of the repeated measures in N and G formats was not found [$F$= 2.194, $df$= 1;122, $p$=.141, $\eta^2$= .018], but it was observed that a significant effect of covariates was related to confidence in the N format [$F$= 12.210, $df$= 1;122, $p$=.001, $\eta^2$= .091] and in G format [$F$= 15.304, $df$= 1;122, $p$=.0001, $\eta^2$= .111].

Additionally, in the Italian sample working without time pressure, no format effect was identified, but the effect of the covariate related to confidence in N [$F$ = 10.579, $df$= 1;343, $p$=.001, $\eta^2$= .030] and G format [$F$= 6.906, $df$= 1;343, $p$=.009, $\eta^2$= .020] was present.

**Table 3.** Linear Regression (backward method). Criterion variable: number of correct responses in G format

| | Block | $R^2$ (SE) | Variables | Beta | t | p | Tolerance | VIF |
|---|---|---|---|---|---|---|---|---|
| Spanish sample without time pressure (S_NTP) n=127 | 1 | .232 (1.178) | PMA VISUO-SPATIAL | .086 | 1.050 | .296 | .908 | 1.101 |
| | | | PMA NUMERICAL | .013 | .163 | .871 | .933 | 1.072 |
| | | | SAS_EXAMINATION | .029 | .317 | .752 | .715 | 1.398 |
| | | | SAS_INTERPRETATION | -.138 | -1.388 | .168 | .613 | 1.631 |
| | | | SAS_HELP | -.015 | -.167 | .868 | .749 | 1.335 |
| | | | SATS_AFFECT | -.115 | -1.156 | .250 | .619 | 1.615 |
| | | | SATS_COMPETENCE | -.027 | -.257 | .797 | .554 | 1.806 |
| | | | SATS_VALUE | .217 | 2.412 | .017* | .754 | 1.326 |
| | | | SATS_DIFFICULTY | .230 | 2.392 | .018* | .660 | 1.515 |
| | | | FG CONFIDENCE | .406 | 4.660 | .0001** | .803 | 1.245 |
| | Last 8 | .243 (1.169) | SATS_VALUE | .222 | 2.780 | .006** | .944 | 1.059 |
| | | | SATS_DIFFICULTY | .189 | 2.374 | .019* | .951 | 1.052 |
| | | | FG CONFIDENCE | .410 | 5.241 | .0001** | .980 | 1.020 |
| Italian sample without time pressure (IT_NTP) n=376 | 1 | .219 (1.206) | PMA VISUO-SPATIAL | .135 | 2.714 | .007* | .872 | 1.146 |
| | | | PMA NUMERICAL | .028 | .565 | .573 | .891 | 1.123 |
| | | | SAS_EXAMINATION | .006 | .104 | .917 | .735 | 1.360 |
| | | | SAS_INTERPRETATION | -.211 | -3.250 | .001* | .514 | 1.945 |
| | | | SAS_HELP | -.017 | -.292 | .770 | .637 | 1.569 |
| | | | SATS_AFFECT | .0001 | -.007 | .994 | .648 | 1.543 |
| | | | SATS_COMPETENCE | -.079 | -1.409 | .160 | .694 | 1.440 |
| | | | SATS_VALUE | .032 | .600 | .549 | .769 | 1.301 |
| | | | SATS_DIFFICULTY | -.070 | -1.189 | .235 | .633 | 1.579 |
| | | | FG CONFIDENCE | .340 | 6.036 | .0001** | .684 | 1.463 |
| | Last 7 | .228 (1.200) | PMA VISUO-SPATIAL | .140 | 2.892 | .004** | .912 | 1.097 |
| | | | SAS_INTERPRETATION | -.206 | -3.992 | .0001** | .804 | 1.243 |
| | | | FG CONFIDENCE | .338 | 6.371 | .0001** | .762 | 1.313 |
| Italian sample with time pressure (IT_TP) n=173 | 1 | .166 (1.151) | PMA VISUO-SPATIAL | -.011 | -.114 | .909 | .846 | 1.183 |
| | | | PMA NUMERICAL | .054 | .546 | .587 | .858 | 1.166 |
| | | | SAS_EXAMINATION | -.025 | -.192 | .848 | .498 | 2.008 |
| | | | SAS_INTERPRETATION | -.025 | -.188 | .851 | .461 | 2.170 |
| | | | SAS_HELP | .028 | .234 | .816 | .569 | 1.757 |
| | | | SATS_AFFECT | .045 | .409 | .683 | .695 | 1.439 |
| | | | SATS_COMPETENCE | .051 | .449 | .654 | .644 | 1.552 |
| | | | SATS_VALUE | .101 | 1.036 | .303 | .868 | 1.153 |
| | | | SATS_DIFFICULTY | -.098 | -.989 | .325 | .847 | 1.180 |
| | | | FN CONFIDENCE | .410 | 3.574 | .001** | .626 | 1.598 |
| | Last 10 | .218 (1.115) | FG CONFIDENCE | .475 | 5.396 | .0001** | 1.000 | 1.000 |

*Note:* * *p<.05;* ** *p<.001*

**Table 4.** Results of mixed design ANCOVA

| Sample | Source | Wilks' Lambda | df | F | p | Eta² | N FORMAT M (SD) | G FORMAT M (SD) |
|---|---|---|---|---|---|---|---|---|
| Spanish sample without time pressure (S_NTP) n=127 | FORMAT | .982 | 1 | 2.194 | .141 | .018 | 2.000 (1.480) | 2.622 (1.344) |
| | FORMAT * SATS VALUE | .997 | 1 | 2.864 | .093 | .023 | | |
| | FORMAT * SATS DIFFICULTY | .998 | 1 | .294 | .589 | .002 | | |
| | FORMAT* FN CONFIDENCE | .909** | 1 | 12.210** | .001 | .091 | | |
| | FORMAT* FG CONFIDENCE | .889** | 1 | 15.304** | .0001 | .111 | | |
| | ERRORS | | 122 | | | | | |
| Italian sample without time pressure (IT_NTP) n=376 | FORMAT | .992 | 1 | 2.766 | .097 | .008 | 1.828 (1.442) | 2.620 (1.370) |
| | FORMAT * PMA VISUO-SPATIAL | .997 | 1 | 1.160 | .282 | .003 | | |
| | FORMAT * SAS INTERPRETATION | .998 | 1 | .852 | .357 | .002 | | |
| | FORMAT* FN CONFIDENCE | .970** | 1 | 10.579** | .001 | .030 | | |
| | FORMAT* FG CONFIDENCE | .980** | 1 | 6.906** | .009 | .020 | | |
| | ERRORS | | 343 | | | | | |
| Italian sample with time pressure (IT_TP) n=173 | FORMAT | .943** | 1 | 8.640** | .004 | .057 | 2.524 (1.487) | 2.940 (1.255) |
| | FORMAT * PMA NUMERICAL | .980 | 1 | 2.849 | .094 | .020 | | |
| | FORMAT* FN CONFIDENCE | .935** | 1 | 9.898** | .002 | .065 | | |
| | FORMAT* FG CONFIDENCE | .979 | 1 | 3.111 | .080 | .021 | | |
| | ERRORS | | 143 | | | | | |

*Note: M = mean; SD= standard deviation; **p < .01*

The significant difference in the means of the correct responses in the N and G formats was observed only in the Italians working under time pressure. They showed a better performance in the G format [$F$=8.640, $df$= 1;143, $p$=.004, $\eta^2$= .057]; moreover, it was confirmed that the significant effect of the covariate was related to confidence in the N format [$F$=9.898, $df$= 1;143, $p$=.002, $\eta^2$= .065].

In summary, the common aspects in the three samples were related to the role of the metacognitive dimension of confidence (Table 4).

## DISCUSSION AND CONCLUSIONS

The aim of this work was to investigate the cognitive processes underpinning probabilistic reasoning in Psychology undergraduates without any statistical expertise in problems in the verbal-numerical and graphical-pictorial formats. In particular, we sought to explore the effect of graphical facilitation (Brase & Hill, 2015; Moro et al., 2011), accounting for the impact of some contextual (e.g., time pressure), metacognitive (e.g., confidence in the correctness of responses), cognitive (e.g., abilities, attitudes) and non-cognitive (e.g., anxiety) dimensions. The study involved Spanish and Italian undergraduates. A subset of the Italian sample worked under time pressure, whereas the remainder of the Italian and Spanish samples worked without time pressure. The current outcomes provide some interesting suggestions that seem to partially clarify the conflicting results present in literature.

Overall, it is possible to conclude that the presentation of problems in the graphical-pictorial format may partially contribute to overcoming the difficulties of inexperienced students in probabilistic reasoning when they work under time pressure. Indeed, this graphical facilitation effect was observed in Italians solving problems under time limits. Moreover, they showed a superior performance in both formats with respect to their colleagues working without time pressure (i.e., Italian and Spanish samples).

Furthermore, in an attempt to identify the variables predicting performance, we highlighted other interesting aspects. In all samples and in both formats, confidence was the stronger predictor of probabilistic reasoning performance. Nevertheless the percentage of explained variance in all regressions was low, which implies that a great part of variance of the criterion is not explained by the identified variables.

Specifically, using performance as the criterion variable in the N format, it was observed that confidence had a significant effect among Spanish students working without time pressure. For the Italians working without time limits, the SAS interpretation and confidence were significant; and for Italian students working under time pressure, the PMA numerical abilities and confidence were significant.

Considering the performance predictors in the G format, it was detected that consistently with the N format, the only significant predictor in all samples was confidence. In the Spanish sample, we also highlighted the effects of SATS value and SATS difficulty (greater than confidence). In the Italian sample working without time pressure, visuo-spatial abilities and anxiety in data interpretation (SAS) were also observed to be significant. Moreover, the unique significant effect of confidence was underlined in the Italian undergraduates working under time pressure.

Subsequently, the application of a mixed design Ancova in relation to each sample allowed us to outline other interesting features. Specifically, the significant predictors in the previous regressions were used as covariates, focusing on the relationships really affecting the performance. Graphical facilitation could be identified only in Italian students working under time pressure. Moreover, this effect appeared to be influenced by confidence in the N format. In other samples (Spanish and Italian students without time pressure) a significant difference in performance across the two formats was not observed.

The results showed that confidence was an important dimension sharing variance in common with the criterion variables of performances, as other studies have shown (Morony et al., 2013; Stankov, 2013; Stankov et al., 2012).

Based on these findings, it is possible to speculate that performance is developed and affected differently in students working with or without time pressure (Ackerman & Lauterman, 2012). Ackerman and Goldsmith (2011) supposed that the mild time pressure gives less freedom for students to apply a learning regulation; conversely, however, it might be also perceived as stimulating (Pintrich, 2003). Indeed, a slight time pressure could support motivated students to apply superior action implementation by disengaging from weakening strategies (Henderson, Gollwitzer, & Oettingen, 2007). In this study, the effects of abilities, attitudes, and anxiety appear to be different under time pressure. Moreover, only with time limits in place there is a significant effect of numerical abilities in the N format. These findings may be related to the literature affirming how undergraduates may change their problem solving strategies on the basis of their abilities, adapting to the requirements of the problem (Wilkinson, Reader, & Payne, 2012).

Consequently, considering the inter-related factors, it is possible to suppose that the usefulness of graphics is affected by some aspects (Dewolf, Van Dooren, Ev Cimen, & Verschaffel, 2014). Generally, the results showed that in two out of three samples, the use of graphical-pictorial representations did not improve students' performance in probabilistic reasoning. These findings may support the hypothesis that there was not a "simple" graphical facilitation effect. Moreover, in the performances, the presence versus absence of time pressure interacted with other dimensions, particularly with the attitudes towards statistics. This is consistent with the work of Tempelaar et al. (2007), who observed that students' attitudes were linked to former knowledge and specific abilities. In this work, when the prediction of performance accuracy in the N and G formats was evaluated, it was only possible to highlight some significant effects of attitudes for the prediction of the G format for the Spanish students who worked without time pressure.

The findings of this research may also be related to the statements of Kellen et al. (2013). They found a direct relationship between the performance in graphical-pictorial format and the level of visuo-spatial abilities. In the present investigation, this relationship was recognised only in Italians working without time pressure, in which significant effects of the anxiety scale in data interpretation and of confidence in the G format were also observed. It was interesting to observe that this sample presented the worst performance in both problem formats compared to the other samples (Italians under time pressure and Spanish students without time pressure). At the descriptive level, they also showed lower levels of visuo-spatial abilities and higher levels of anxiety in data interpretation (compared with the other two samples).

Taking account of these dimensions, our findings might be also evaluated in relation to the Stanovich's model of Dual-Process (e.g. Evans & Stanovich, 2013) that defines the close relation between intuitive and reflective data processing. The probabilistic tasks given to our undergraduates without any statistical expertise, could be solved by the use of Type I processing, in which is used non-normative, non-conscious, instinctive, associative and automatic knowledge. This Type I of processing could appear mainly in relation to the verbal -numerical format of problem; this aspect may be due to the similarity to the classical format of mathematical problems presentation at school. However, coherently, a stronger role of the processing Type II might be implied in the graphical -pictorial format of problems. In this format, as consistently the literature shows, we can observe an effect of facilitation and the more specific significant effect of metacognitive dimension of confidence (e.g. Stankov, 2013; Stankov et al., 2012; Tempelaar, 2009; Tempelaar et al., 2007; Thompson et al., 2011). The application of Type II of processing seems to be enhanced in the G format, in where problems have a minor similarity to the classical probabilistic problems solved by students at high school (Horton & Aykin, 2005). Moreover, the effect of graphical facilitation seems to be supported by the presence of a mild time pressure, that don't allow to students to apply cognitive ruminations on their abilities, reducing the attitudes and anxiety effects (Ackerman & Lauterman, 2012; Henderson, Gollwitzer, & Oettingen, 2007; Maule et al., 2000).

Our outcomes could seem a little evidence to agree with the *nested sets* approach to probabilistic reasoning (Sirota, Juanchich, & Hagmayer, 2014), supporting the relevance of thinking dispositions and metacognition in the solution of probabilistic problems, both in verbal-numerical than in graphical-pictorial formats (Stanovich, 1999; Stanovich & West, 1998; Thompson et al., 2011; 2014). Furthermore, in our work, when the students work without time limits, we confirmed the *nested sets* individuated by Sirota et al. (Sirota, Kostovičová, & Juanchich, 2014); but, when the same problems were administered in a mild time pressure, the effect of graphical facilitation might appear stronger and significative than in other conditions, as potential results of application of a deeper and analytical cognitive data processing (Stanovich, 1999).

Our findings could be coherent with the statements of Sirota et al. (Sirota, Kostovičová, & Juanchich, 2014); they underline that when the problem structure yet enlightens the *nested sets* data, in this case the graphical format don't support the reasoning more than verbal-numerical format (as we observed in our data when probabilistic problems were solved without time pressure). Nevertheless, in our work we observed a significant effect of graphical facilitation in time pressure. This result could be explained by the enhancement of commitment and by an encouragement to apply a deeper data processing in presence of time limits (Maule et al., 2000), which could amplify the reflection on *nested sets*, the application of Type II processing in both formats, but especially in the graphical-pictorial.

Moreover our results could be a little evidence of the role of individual differences in the probabilistic reasoning, in which attitudes and abilities (as numeracy and visuo-spatial skills) could have an influence in the application of reasoning (Sirota & Juanchich, 2011). Our finding moreover might be considered coherent in the light of *nested sets* approach, with the *problem interpretation hypothesis* (Sirota, Kostovičová, & Vallée-Tourangeau, 2015), supporting the idea that the problem presentation structure might recall in memory specific information related to mathematical operators useful to solve the problem (e.g., Kintsch & Greeno, 1985); from this aspect could derive the facilitation effect. Specifically, an adequate mental representation of the problem might trigger the suitable information, outlines and mathematical procedures, improving in this way the probabilistic reasoning (e.g., Sirota et al., 2015).

Then it is possible to presume that the graphical-pictorial format was not facilitating probabilistic reasoning *per se*. It is reasonable to speculate that undergraduates without statistical knowledge could sometimes fail to appreciate a graphic because they cannot understand it or recognise the data it is supposed to describe (Rundgren & Tibell, 2010). Moreover, it was useful to observe that many students in this study declared in the open responses that they did not understand the structure of the graphics and preferred reading the text. This preference could also be associated with the students' learning and cognitive styles, which may carry a preference for the verbal-numerical format of problem presentation (Kirby, Moore, & Schofield, 1988; Sternberg & Grigorenko, 1997). This preference could also be related to the common cultural habits in Western and European countries, where the classical learning and teaching approach to numerical problems does not often include graphical-pictorial representations (Horton & Aykin, 2005).

These findings leave many questions unanswered related to the specificities of Italian and Spanish undergraduates, who use different curricula. Moreover, the 2012 PISA reports highlighted some dissimilarities between Italian and Spanish learners. In this study, students' performance in Italy and Spain was similar in mathematics, reading, and science, but appeared to be different in problem solving (http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm). Indeed, Italians showed a good performance on interactive tasks and knowledge acquisition problems; Spanish students, however, showed a similarly good performance on interactive tasks, but a weaker performance in knowledge acquisition tasks (http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n38-(eng)-final.pdf). It is possible to speculate that this distinction may indicate real differences in educational systems, characterize and define respective strengths and weaknesses in each country.

It is also important to reflect on the limits of this work. It could be useful to assess students' mathematical and statistical self-efficacy in the protocol in order to control their effect on performance and their relationships with other variables. Another important psychological dimension to investigate could be the locus of control, in order to assess the role of causal attributions and their associations with cognitive and metacognitive dimensions (especially confidence). Investigating cognitive styles and learning styles could also undoubtedly be advantageous in order to explain differences in performance in the two formats (Blais, Thompson, & Baranski, 2005). They may exert an important influence on the supposed effect of graphical facilitation, interacting with the other dimensions. Coherently, in order to better understand the performance, it could be worthwhile to analyse the cognitive processes applied in problem solving compared with the probabilistic reasoning carried out (Lorenzo, 2005; Polaki, 2005). Another crucial aspect may be related to the assessment of social desirability, which could exert an influence on the self-reported evaluations of attitudes and anxiety (Fastame & Penna, 2012).

Finally, it is possible to outline some of this work's strong points. One is its evaluation of the effects of graphical facilitation (Brase & Hill, 2015; Moro et al., 2011), comparing similar problems solved by the same inexperienced students, controlling the influence of cognitive, metacognitive and non-cognitive variables that have been classically identified in previous studies, under conditions of the presence versus absence of time pressure (Guàrdia-Olmos et al., 2006; Morony et al., 2013). The simultaneous evaluation of these dimensions constitutes an innovation in this work. Further, the assessment of the influence exerted by visuo-spatial skills has generally been marginalized. Nevertheless, it appears to be noteworthy in shedding light on graphical facilitation in problem solving (Kellen et al., 2013; Zhu & Gigerenzer, 2006).

In conclusion, this work provides some stimulating outcomes in relation to the way in which psychology undergraduates lacking statistical knowledge can handle probabilistic problems. The findings may be useful to shed light on the usefulness of graphical-pictorial representations in probabilistic reasoning.

These outputs highlight the utility of a broader approach to the understanding of performance in verbal-numerical and graphical-pictorial formats and of effect of graphical facilitation. The use of the frequentist approach with problems in simple formulations could support probabilistic reasoning in undergraduates without statistical knowledge. Kellen et al. (2013) stated that this fact was apparent in the N format in the case of low visuo-spatial abilities and in the G format of problem presentation in the case of high visuo-spatial abilities.

These results may provide some implications for empirical matters and help to outline specific intervention programs useful to support and enhance this reasoning, which is crucial in statistics.

## REFERENCES

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: on screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18-32. doi: http://dx.doi.org/10.1037/a0022086

Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior*, *28*(5), 1816–1828. doi:10.1016/j.chb.2012.04.023

Agus, M., Peró-Cebollero, M., Guàrdia-Olmos, J., & Penna, M. P. (2014). Graphical statistics as an option for the improvement of learning in Psychology. In & R. G. K. Makar, B. de Sousa (Ed.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014)*. Flagstaff - AZ - USA: Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C256_AGUS.pdf

Agus, M., Peró-Cebollero, M., Penna, M., & Guàrdia-Olmos, J. (2015). Towards the development of problems comparing verbal-numerical and graphical formats in statistical reasoning. *Quality & Quantity*, *49*(2), 691–709. http://doi.org/10.1007/s11135-014-0018-7

Agasisti, T., & Cordero-Ferrera, J. M. (2013). Educational disparities across regions: A multilevel analysis for Italy and Spain. *Journal of Policy Modeling*, *35*(6), 1079–1102. doi:10.1016/j.jpolmod.2013.07.002

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983-998.doi: http://dx.doi.org/10.1037/0278-7393.33.6.983

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*(4), 584–600. doi:http://dx.doi.org/10.1037/0096-3445.133.4.584

Ben-Zvi, D., & Garfield, J. B. (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Kluwer Academic Pub.

Blais, A.-R., Thompson, M. M., & Baranski, J. V. (2005). Individual differences in decision processing and confidence judgments in comparative judgment tasks: The role of cognitive styles. *Personality and Individual Differences*, *38*(7), 1701–1713. doi:10.1016/j.paid.2004.11.004

Brase, G. L. (2000). "Bugs" built into the system: *Learning and Individual Differences*, *12*(4), 391–409. doi:10.1016/S1041-6080(02)00048-1

Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, *23*(3), 369–381. doi:10.1002/acp.1460

Brase, G. L., & Hill, W. T. (2015). Good fences make for good neighbors but bad science: a review of what improves Bayesian reasoning and why. *Frontiers in Psychology, 6,* 340. doi:10.3389/fpsyg.2015.00340

Britt, D. W., & Chen, Y.-C. (2013). Increasing the capacity of conceptual diagrams to embrace contextual complexity. *Quality & Quantity*, *47*(1), 567–576. doi:10.1007/s11135-011-9479-0

Carmona-Màrquez, J. C. (2004). Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística. *Statistics Education Research Journal*, *3*(1), 5–28. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ3(1).pdf#page=8

Chiesi, F., & Primi, C. (2009a). Un modello sul rendimento nelle materie quantitative degli studenti di psicologia. *Giornale Italiano Di Psicologia*, *36*(1), 161–184.

Chiesi, F., & Primi, C. (2009b). Assessing statistics attitudes among college students: Psychometric properties of the Italian version of the Survey of Attitudes toward

Statistics (SATS). *Learning and Individual Differences*, *19*(2), 309–313. doi:10.1016/j.lindif.2008.10.008

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal*, *9*(1), 6–26. Retrieved from http://iase-web.org/documents/SERJ/SERJ9(1)_Chiesi_Primi.pdf

Chiesi, F., Primi, C., & Carmona, J. (2011). Measuring Statistics Anxiety. Cross-Country Validity of the Statistical Anxiety Scale (SAS). *Journal of Psychoeducational Assessment*, *29*(6), 559–569. doi:10.1177/0734282911404985

Colet, A. V, Seva, U. L., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema*, *20*(1), 174–180.

Colom, R., Contreras, M. J., Botella, J., & Santacreu, J. (2002). Vehicles of spatial ability. *Personality and Individual Differences*, *32*(5), 903–912. doi:10.1016/S0191-8869(01)00095-2

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1–73. doi:10.1016/0010-0277(95)00664-8

Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, *29*(3), 89–93. doi:10.1111/j.1467-9639.2007.00267.x

delMas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy*. Auckland, New Zealand: University of Auckland. Retrieved from https://www.causeweb.org/artist/articles/SRTL4_ARTIST.pdf

Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *The Journal of Experimental Education*, *82*(1), 103–120. doi:10.1080/00220973.2012.745468

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. http://doi.org/10.1177/1745691612460685

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology*, *63*(5), 892–909. http://doi.org/10.1080/17470210903111821

Evans, J. S. B. T., Thompson, V., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.00398

Fastame, M. C., & Penna, M. P. (2012). Does social desirability confound the assessment of self-reported measures of well-being and metacognitive efficiency in young and older adults? *Clinical Gerontologist*, *35*(3), 239–256. doi:10.1080/07317115.2012.660411

Fernández-Aguirre, K., Garín-Martín, M. A., & Modroño-Herrán, J. I. (2014). Visual displays: analytical study and applications to graphs and real data. *Quality & Quantity*, *48*(4), 2209–2224. doi:10.1007/s11135-013-9887-4

Fischer, S. C., Hickey, D. T., Pellegrino, J. W., & Law, D. J. (1994). Strategic processing in dynamic spatial reasoning tasks. *Learning and Individual Differences*, *6*(1), 65–105. doi:10.1016/1041-6080(94)90015-9

Galli, S., Chiesi, F., & Primi, C. (2011). Measuring mathematical ability needed for "non-mathematical" majors: The construction of a scale applying IRT and differential item functioning across educational contexts. *Learning and Individual Differences*, *21*(4), 392–402. doi:10.1016/j.lindif.2011.04.005

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2011). Cómo favorecer la comprensión y la comunicación de los riesgos sobre la salud. *Psicothema*, *23*(4), 599–605.

Garfield, J., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, *67*(1), 1–12. doi:10.1111/j.1751-5823.1999.tb00377.x

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. http://doi.org/http://dx.doi.org/10.1037/0033-295X.102.4.684

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528. doi: http://dx.doi.org/10.1037/0033-295X.98.4.506

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, *78*(3), 247–276. http://doi.org/10.1016/S0010-0277(00)00133-5

Guàrdia-Olmos, J., Freixa-Blanxart, M., Peró-Cebollero, M., Turbany, J., Cosculluela, A., Barrios, M., & Rifà, X. (2006). Factors related to the academic performance of students in the statistics course in psychology. *Quality & Quantity*, *40*(4), 661–674. doi:10.1007/s11135-005-2072-7

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences*, *1*(2), 78–82. doi:10.1016/S1364-6613(97)01014-0

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689. doi:10.1037/0022-0663.91.4.684

Henderson, M. D., Gollwitzer, P. M., & Oettingen, G. (2007). Implementation intentions and disengagement from a failing course of action. *Journal of Behavioral Decision Making*, *20*(1), 81–102. doi:10.1002/bdm.553

Hoffman, B. (2010). "I think I can, but I'm afraid to try": The role of self-efficacy beliefs and mathematics anxiety in mathematics problem-solving efficiency. *Learning and Individual Differences*, *20*(3), 276–283. http://doi.org/10.1016/j.lindif.2010.02.001

Horton, W., & Aykin, N. (2005). Graphics: The not quite universal language. In N. Aykin (Ed.), *Usability and internationalization of information technology* (pp. 157–188). CRC Press.

Jackson, S. A., & Kleitman, S. (2013). Individual differences in decision-making and confidence: capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9*(1), 25–49. doi:10.1007/s11409-013-9110-y

Johnson-Laird, P. N. (1983). *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Harvard University Press.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, *19*(4), 201–214. http://doi.org/10.1016/j.tics.2015.02.006

Johnson, E. D., & Tubau, E. (2013). Words, numbers, & numeracy: Diminishing individual differences in Bayesian reasoning. *Learning and Individual Differences*, *28*, 34–40. doi:10.1016/j.lindif.2013.09.004

Juanchich, M., & Sirota, M. (2015). How to improve people's interpretation of probabilities of precipitation. *Journal of Risk Research*, 1–17. http://doi.org/10.1080/13669877.2014.983945

Kellen, V., Chan, S., & Fang, X. (2006). Individual Differences in Spatial Abilities and the Visualization of Conditional Probabilities. Retrieved from http://www.kellen.net/visualization_wp.htm

Kellen, V., Chan, S., & Fang, X. (2007). Facilitating Conditional Probability Problems with Visuals. In J. Jacko (Ed.), *Human-Computer Interaction. Interaction Platforms and Techniques* (Vol. 4551, pp. 63–71). Berlin / Heidelberg: Springer. doi:10.1007/978-3-540-73107-8_8

Kellen, V., Chan, S., & Fang, X. (2013). Improving user performance in conditional probability problems with computer-generated diagrams. In *Human-Computer Interaction. Users and Contexts of Use* (pp. 183–192). Berlin / Heidelberg: Springer.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*(1), 109. http://doi.org/http://dx.doi.org/10.1037/0033-295X.92.1.109

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, *13*(2), 169–184. doi:10.1016/0361-476X(88)90017-3

Klaczynski, P. A. (2014). Heuristics and biases: interactions among numeracy, ability, and reflectiveness predict normative responding. *Frontiers in Psychology*, *5*, 665. http://doi.org/10.3389/fpsyg.2014.00665

Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, *30*(3), 363–371. doi:10.3758/BF03194937

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *25*(1), 108–125. doi:http://dx.doi.org/10.1037/h0078792

Lorenzo, M. (2005). The development, implementation, and evaluation of a problem solving heuristic. *International Journal of Science and Mathematics Education*, *3*(1), 33–58. doi:10.1007/s10763-004-8359-7

Mandel, D. R. (2014a). The psychology of Bayesian reasoning. *Frontiers in Psychology, 5,* 1144. doi:10.3389/fpsyg.2014.01144

Mandel, D. R. (2014b). Visual representation of rational belief revision: another look at the Sleeping Beauty problem. *Frontiers in Psychology, 5,* 1232. doi:10.3389/fpsyg.2014.01232

Mandel, D. R. (2015). Instruction in Information Structuring Improves Bayesian Judgment in Intelligence Analysts. *Frontiers in Psychology, 6,* 387 http://doi.org/10.3389/fpsyg.2015.00387

Maule, A. J., Hockey, G. R. J., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica*, *104*(3), 283–301. doi:10.1016/S0001-6918(00)00033-0

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. doi: http://dx.doi.org/10.1037/0033-295X.115.2.502

Moro, R., & Bodanza, G. A. (2010). El debate acerca del efecto facilitador en problemas de probabilidad condicional: ¿Un caso de experimentación crucial? *Interdisciplinaria*, *27*(1), 163–174.

Moro, R., Bodanza, G. A., & Freidin, E. (2011). Sets or frequencies? How to help people solve conditional probability problems. *Journal of Cognitive Psychology, 23*(7), 843–857. doi:10.1080/20445911.2011.579072

Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research*, *58*, 79–96. doi:10.1016/j.ijer.2012.11.002

Oaksford, M. (2015). Imaging deductive reasoning and the new paradigm. *Frontiers in Human Neuroscience*, *9*(FEB). http://doi.org/10.3389/fnhum.2015.00101

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667–686. doi: http://dx.doi.org/10.1037/0022-0663.95.4.667

Polaki, M. V. (2005). Dealing with compound events. In G. A. Jones (Ed.), *Exploring Probability in School* (pp. 191–214). New York: Springer. doi:10.1007/0-387-24530-8_9

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*(2), 258–276. doi:10.1016/j.actpsy.2007.05.004

Rundgren, C.-J., & Tibell, L. A. E. (2010). Critical features of visualizations of transport through the cell membrane – An empirical study of upper secondary and tertiary students' meaning – Making of a still image and an animation. *International Journal of Science and Mathematics Education*, *8*(2), 223–246. doi:10.1007/s10763-009-9171-1

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The Development and Validation of the Survey of Attitudes toward Statistics. Educational and Psychological Measurement, 55(5), 868–875. doi:10.1177/0013164495055005022

Sirota, M., & Juanchich, M. (2011). Role of numeracy and cognitive reflection in Bayesian reasoning with natural frequencies. *Studia Psychologica, 53*(2), 151–161. Retrieved from http://yadda.icm.edu.pl/cejsh/element/bwmeta1.element.defbace5-44e3-3d1b-8464-09dde16918e3

Sirota, M., Juanchich, M., & Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review, 21*(1), 198–204. http://doi.org/10.3758/s13423-013-0464-6

Sirota, M., Kostovičová, L., & Juanchich, M. (2014). The effect of iconicity of visual displays on statistical reasoning: evidence in favor of the null hypothesis. *Psychonomic Bulletin & Review, 21*(4), 961–968. http://doi.org/10.3758/s13423-013-0555-4

Sirota, M., Kostovičová, L., & Vallée-Tourangeau, F. (2015). Now you Bayes, now you don't: effects of set-problem and frequency-format mental representations on statistical reasoning. *Psychonomic Bulletin & Review*, 1–9. http://doi.org/10.3758/s13423-015-0810-y

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296–309. doi:10.1016/S0749-5978(03)00021-9

Sorge, C., & Schau, C. (2002). Impact of engineering students' attitudes on achievement in statistics: A structural model. In *Annual meeting of the American Educational Research Association*. (Vol. New Orleans). Retrieved from http://evaluationandstatistics.com/AERA2002.pdf

Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences*, *55*(7), 727–732. doi:10.1016/j.paid.2013.07.006

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, *22*(6), 747–758. doi:10.1016/j.lindif.2012.05.013

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In J. St. B. T. Evans & K. Frankish (Ed.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford, England: Oxford University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford, England: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? *The Behavioral and Brain Sciences*, *23*(5), 645–665; discussion 665–726.

Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style? *American Psychologist,* 52(7), 700–712. doi:10.1037/0003-066X.52.7.700

Tempelaar, D. T. (2009). The Role of Self-theories of Intelligence and Self-perceived Metacognitive Knowledge, Skills, and Attitudes, in Learning Statistics. In *Fifth Global SELF International Biennial Conference. Enabling Human Potential: The Centrality of Self and Identity Constructs*. (Vol. 13–15 Ja). Retrieved from http://www.self.ox.ac.uk/documents/Tempelaar.pdf

Tempelaar, D. T., Van Der Loeff, S. S., & Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal*, *6*(2), 78–102.

Thompson, V. A., Prowse, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–40. http://doi.org/10.1016/j.cogpsych.2011.06.001

Thurstone, L. L., & Thurstone, T. G. (1981). *PMA: abilità mentali primarie: manuale di istruzioni Livello intermedio (11-17)(Batteria fattoriale delle abilità mentali primarie)*. Firenze: Organizzazioni Speciali.

Thurstone, L. L., & Thurstone, T. G. (1987). *TEA: tests de aptitudes escolares [niveles 1, 2 y 3]: manual*. Madrid: Tea.

Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioral Science*, *32*(1), 40–48. doi: http://dx.doi.org/10.1037/h0087099

Tubau, E. (2008). Enhancing probabilistic reasoning: The role of causal graphs, statistical format and numerical skills. *Learning and Individual Differences*, *18*(2), 187–196. doi:10.1016/j.lindif.2007.08.006

Wilkinson, S. C., Reader, W., & Payne, S. J. (2012). Adaptive browsing: Sensitivity to time pressure and task difficulty. *International Journal of Human-Computer Studies*, *70*(1), 14–25. doi:10.1016/j.ijhcs.2011.08.003

Yamagishi, K. (2003). Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)*, *50*(2), 97–106. doi:10.1026//1618-3169.50.2.97

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287–308. doi:10.1016/j.cognition.2004.12.003

Pisa Reports 2012. (n.d.-a). Retrieved from http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n38-(eng)-final.pdf

Pisa Reports 2012. (n.d.-b). Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm

◈◇◈