anales de psicología, 2016, vol. 32, nº 3 (octubre), 893-898
http://dx.doi.org/10.6018/analesps.32.3.224401

ISSN edición impresa: 0212-9728. ISSN edición web (http://revistas.um.es/analesps): 1695-2294

JOLer: A Java standalone application for simulating the Weaver & Kelemen's judgment of learning (JOL) model

Marcos Ruiz* and Cristóbal Arroyo

Universidad Nacional de Educación a Distancia, Madrid (Spain).

Título: JOLer: Una aplicación Java de escritorio para simular el modelo de juicios de aprendizaje de Weaver y Kelemen.
Resumen: Para calcular la precisión de los juicios de aprendizaje (JJAA) en metamemoria los investigadores tienen que estimar en qué medida los juicios de un participante se ajustan a su rendimiento en una prueba de memoria. La precisión absoluta o *calibración* es la correspondencia media entre JA y rendimiento en memoria. La precisión relativa de metamemoria o *resolución* nos dice el grado de sensibilidad de un participante respecto a un diferencial de recuperabilidad entre dos ítems estudiados. Por desgracia, los factores que alteran la calibración y la resolución con frecuencia cambian también la distribución de JJAA a lo largo de la escala de juicio. El problema de estos efectos sobre la distribución de JJAA es que pueden dar lugar a una estimación de resolución distorsionada debido al modo en que se calcula el estimador habitual. *JOLer* simula el comportamiento de unos participantes en un procedimiento típico de metamemoria. La aplicación se presenta como una herramienta para investigadores de la metamemoria: ofrece la oportunidad de comprobar si, manteniendo los parámetros de calibración pero cambiando la distribución de JJAA entre condiciones, se obtendría una resolución estimada distinta (y en cierto grado artificial).
Palabras clave: metamemoria; metacognición; juicio de aprendizaje; JOL; correlación gamma; simulación; Java.

Abstract: To assess *judgment of learning* (JOL) accuracy in metamemory, researchers have to measure how much the metamemory judgments adjust to the participant's memory-test performance. Absolute accuracy or *calibration* is the average correspondence between JOL and memory performance. Metamemory relative accuracy or *resolution* is a measure of how sensitive a participant is to the differential recallability between two studied items. Unfortunately, factors altering both calibration and resolution very often change also the distribution of JOL on the available scale for judgment. The problem with these effects on JOL distribution is that they could yield an altered resolution estimation due to the way in which its usual estimate is computed. *JOLer* simulates the behavior of participants in a typical metamemory procedure. The application is offered as a tool for metamemory researchers: it affords the opportunity to check whether, maintaining calibration parameters but changing JOL distributions between conditions, a different (and somewhat spurious) resolution estimate would be obtained.
Key words: metamemory; metacognition; judgment of learning; JOL; gamma correlation; simulation; Java.

Metamemory Judgments

For a few decades now there has been an increasing interest in the study of metacognitive processes (for review see Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009; Reder, 1996; Ruiz, 2004). From the seminal works by Hart (1965, 1967) on the feeling of knowing (FOK), by Arbuckle & Cuddy (1969) on judgments of learning (JOL), or the ones by Underwood (1966) on ease of learning (EOL) and Brown & McNeill (1966) on the tip-of-the-tongue (TOT) phenomenon, the field has been populated by a considerable amount of contributions on theories (e.g., Kelley & Jacoby, 1996; Koriat, 1993, 1997; Metcalfe, Schwartz, and Joaquim, 1993; Sikström & Jönsson, 2005), procedures (e.g., Glenberg & Epstein, 1987; Glucksberg & McCloskey, 1981; Lovelace, 1984; Nelson & Narens, 1990), and data (e.g., Koriat & Levy-Sadot, 2001; Son & Metcalfe, 2005; Nelson, Leonesio, Landwehr, and Narens, 1986; Nelson & Narens, 1980; Shanks & Serra, 2014; Vesonder & Voss, 1985).

Among the issues addressed by metamemory researchers, perhaps the JOL accuracy is one of the most popular. In a JOL experiment participants are requested, after studying some target material, to make an estimate as to how likely they expect to answer successfully a question about that material (Arbuckle & Cuddy, 1969; Lovelace, 1984). Typically a set of unrelated word-pairs is presented for memorizing. After a variable delay, they are asked for the probability (usually in a scale from 0 to 100) with which they expect to answer the response-word of a pair when presented with its stimulus-word (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Metcalfe et al., 1993).

To assess JOL accuracy researchers have to measure the degree to which the metamemory judgments adjust to the participant's memory-test performance. Interestingly, there are two types of JOL accuracy that have to be considered. Absolute accuracy or *calibration* is the average correspondence between JOL and memory performance (e.g., Finn & Metcalfe, 2014). Metamemory and memory are usually measured on the same scale, as when estimated probability of recall is required at JOL and percentage of correct recall is measured at the final memory test. The simplest index of calibration is the signed difference between the JOL estimates and the memory performance. Researchers make also use of calibration functions or curves, in which the mean recall level is computed for every JOL estimate or bin of estimates. Whatever the way we measure absolute calibration, a mean JOL over the mean memory performance would be indicative of a bias towards overconfidence, while a JOL below their memory counterpart would be an index of underconfidence (e.g., Finn & Metcalfe, 2007; Koriat, Ma'ayan, Sheffer, & Bjork, 2006; Koriat & Nussinson, 2009; Nelson & Dunlosky, 1991).

Another type of JOL accuracy is the relative accuracy or *resolution* (e.g., e.g., Finn & Metcalfe, 2014 ; Koriat, 1993). Metamemory resolution is a measure of how sensitive a

*** Dirección para correspondencia [Correspondence address]:**
Marcos Ruiz. Departamento Psicología Básica I. Facultad de Psicología. UNED. C/ Juan del Rosal, 10. 28040. Madrid (Spain).
E-mail: marcos@psi.uned.es

paticipant is to any differential accessibility or recallability between two studied items. Notice that there has not to be any difference in regards to the JOL task as compared to the participant's task for the calibration measure. But when an individual's metamemory resolution is assessed, we wonder if an item with a higher JOL than another one has a higher recall probability as well (e.g., Liberman & Tversky, 1993). So resolution should be viewed as the power of the lens with which somebody monitors his/her own knowledge.

Notice that the JOL resolution measure should be an association index of the difference on an ordinal scale between two JOLs with the corresponding hit/failure memory performances. Yet, after a much-celebrated paper by Nelson (1984) on a few of candiate association indices, very often the Goodman-Kruskal gamma correlation between JOL and memory performance is computed for every participant in an experiment (Goodman & Kruskal, 1954; Nelson, 1984).

There are a number of factors that change simultaneously both calibration and resolution. For instance, the so-called delayed-JOL effect appears when participants make their JOL after a certain delay from study. Yet, JOL resolution as much as calibration are greately improved for delayed JOL, as compared to the JOL made immediately after study (Dunlosky & Nelson, 1992, 1994; Nelson & Dunlosky, 1991; Weaver, Terrell, Krug, & Kelemen, 2008).

A more complex pattern of metamemory-accuracy change appears in the underconfidence with practice (UWP) effect (Koriat, Sheffer, & Ma'ayan, 2002). This phenomenon appears when a few study-JOL-test cycles are run for the same participant. As participant goes from a first to a second study-JOL-test cycle, the memory performance increases while the mean JOL decreases from an overconfidence for the first JOL block to an underconfidence for the second and succesive JOL blocks. Most interestingly, the less confident the participant is, the higher the resolution (see also, Koriat, Ma'ayan, Sheffer, & Bjork, 2006; Koriat et al., 2002; Finn & Metcalfe, 2007, 2014). In summary, calibration and resolution are two measurable dimensions of metacognitive judgments that can exhibit associate as much as dissociate behaviors.

Of special interest here is that the factors altering both calibration and resolution very often change also the distribution of JOL on the available scale for judgment. For instance, Dunlosky & Nelson (1994) reported that when the JOL for an item is given immediately after its study presentation most of the judgment responses on a scale from 0 to 100 were between 20 and 80, so that the extreme JOLs (i.e., "definitely I will remember" and "definitely I will not remember") were fairly scarce. For the delayed-JOL condition the situation was clearly reversed, as most of the JOL responses were extreme values (see also Koriat & Goldsmith, 1996, for a computational modeling of this changes on JOL distribution). Also, a change in the JOL pattern has been reported for the UWP effect. Certainly, the definitely-I-will-remember responses have been reported to be relatively more frequent for the second JOL block and beyond in the study-JOL-test cicles (Koriat et al., 2002).

The problem with these effects on JOL distribution is that they could yield an inflated resolution estimation due to the way in which gamma is computed. Let's assume for a moment a participant with a perfect calibration function. For this participant a JOL of 10 on a 0-100 scale would mean that just 10% of the items that received this JOL will be correctly recalled. The same simple computations have to done for items with JOL= 40, 50, or 90. The point here is that most of the items with JOL=90 will be correctly recalled while just a few of the JOL=10 will be recalled. So the gamma will benefit much from the fact that within the set of both JOLs (10+90) a great proportion of contrasting pairs will be correctly ordered as non-recalled$_{10}$/recalled$_{90}$. This would not be the case for the set of JOL=40 and 50. Here similar proportions of item in each one of the JOL category will be correctly recalled (40% and 50%, respectively, due to the perfect calibration). Therefore, the proportion of contrasting pairs of items correctly ordered within the set of these JOL categories will likely be relatively smaller. In other words, the way in which gamma is calculated could yield that a factor altering just the JOL distribution could spuriously alter the gamma as resolution estimator.

Due to these and other considerations there have been some criticisms regarding the use of gamma as a resolution measure (e.g., Benjamin & Díaz, 2008; González & Nelson, 1996; Masson & Rotello, 2009; Muruyama, Sakaki, Yan, & Smith, 2014). In fact, many authors use alternative estimates for metacognitive resolution (e.g., Arnold, Higham, & Martín-Luengo, 2013; Luna, Higham, Martín-Luengo, 2011). Even so, very often the Goodman-Kruskal gamma correlation between JOL and memory performance is computed for every participant as one of the best choices for individual estimate of metacognitive resolution (e.g., Metcalfe & Finn, 2008; Pyc, Rawson, & Aschenbrenner, 2014; Serra & Ariel, 2014; Sundqvist, Todorov, Kubik, & Jönsson, 2012; for a metanalysis see Rhodes & Tauber, 2011). As a consequence, a crucial question for researchers is to take into account the effects on gamma due to changes of JOL distribution between conditions. We present here a software tool to find out the presence of these effects within collected data. The program logic was forwarded by Weaver & Kelemen (1997, 2003) but to our knowledge it has not been applied by other researchers, perhaps due to the lack of a ready-to-use implementation.

The Weaver & Kelemen's (1997) simulation strategy

To deal with the problem of gamma inflation due to change in judgments distributions, Weaver and Kelemen (1997) suggested that we could simulate JOLs based on the value of empirically obtained distributions and calibration

parameters. With the simulations we could check whether, maintaining the calibration parameters but changing the JOL distribution between conditions, a higher gamma would be obtained. Were this the case, we should not interpret the whole increase in our experimental gamma as resolution improvement from one experimental condition to other. As complementary information, we could also check how much would have changed the resolution from one experimental condition to the other due to changes in the calibration such as those obtained in our empirical data, but without the accompanying change in judgments distribution.

In the JOL simulation model implemented by Weaver & Kelemen (1997) there are two starting sets of data for every one of our experimental conditions: the JOL distribution and the JOL calibration curve. From these data a set of pairings between a JOL and a memory test can be generated by the model. And the gamma for that set of trials can be computed as the gamma for a simulated participant. An example could be in order.

Suppose that we want to check whether our experimental effect on JOL resolution measured as gamma estimates has been artifactually produced by a change in the JOL distribution. Our empirically collected relative frequencies for the control condition could have been .07, .15, .22, .27, .16, and .13 for corresponding JOLs from 0 to 100 in steps of 20 units (0%, 20%, …, 100%, from sure not to sure yes I will remember). Of course, we have also a different distribution of empirically collected relative frequencies in another experimental condition. The first step in the model is to generate a JOL for a trial selecting a random number in the range 0-1 out of a uniform distribution (say .128). And the JOL corresponding to that bin on the JOL scale will be taken as the JOL response for that trial (20 in our example, as $.07<.128\leq(.07+.15)$).

The second step in the model is to produce a response for the memory test of the item for which we know the JOL generated. For that response to be produced we will draw on the calibration function, instead of the JOL distribution function used for the JOL generation. A new random number in the inclusive range 0-1 is generated from a uniform distribution. If the number is bigger than the conditional probability of recall associated with that JOL in the calibration curve, a correct recall is produced, otherwise a recall failure. For instance, if in our calibration curve a 30% of correct recall is associated with the JOL=20 of our example (a case of underconfidence), with a .432 random number for the recall generator, a correct recall should be recorded as the response for that trial.

Notice that this basic JOL+recall simulation mechanism should be repeated for as many simulated trials as we want per participant. Yet, once a participant has been simulated, we can compute the corresponding gamma. And for every one of our experimental conditions we will have a number of participants simulated. That way, we can define an experimental condition in terms of both its JOL distribution and its calibration curve.

The Java application

JOLer is a Java standalone application aimed to run a huge number of Monte Carlo simulations of metacognitive experiments. It has been developed to check whether the effect of some factor on the JOL distribution is unduly altering the gammas as resolution estimates. It is a cross-platform application as it works on any system running the Oracle Java Virtual Machine (7.0 or higher). The menu-driven (in English or Spanish) user's interface gives us the oportunity to easily define the value of global parameters for up to 10 experimental conditions, with up to 200 participants per condition. For a simulation, between 10 and 200 items per participants can be defined (i.e., memory list-length) and between 2 and 15 bins on the JOL scale, a range covering by far the needs for most of the experimental settings.

Once you have defined the global parameters, the specific settings per condition should be defined. As we said before, an experimental condition is defined on both its JOL distribution and its calibration curve. The application predefines some default values, appropriate for the global simulation parameters. However, for the simulation to be properly used, the experimenter is requested to define a relative frequency for every JOL bin as a promille score. The scores for the calibration curve (i.e., the probability of correct recall associated with every bin) should also be entered as promilles. It should be pointed out that the control condition in a simulation (condition 0 in the JOLer user's interface) should ideally be defined with the JOL distribution and the calibration curve empirically yielded by some real experiment.

Each one of the remaining defined conditions will be compared with the control condition, to check whether the different simulation settings yield different gammas. The point of interest here is that if a condition with the same calibration curve but different JOL distribution than the control condition gives rise to an increased average gamma, this should be considered a spurious resolution improvement. As pointed out by Weaver & Kelemen (1997), given that the number of experiment replications can be arbitrarily increased, the usual ANOVA is pointless here. As they do, *JOLer* gives you information about the median test between the control and any other of the experimental conditions.

There are two special features of *JOLer* that are worth some additional considerations. We should begin noticing that as the model has been detailed before there is no room for random variability other than the one associated with the random number generator for every JOL-recall pair. It should be welcome here to afford for some between-participant variability in regards to JOL distribution and to calibration curve (see Note 2 by Weaver & Kelemen, 1997). A typical participant would be in a certain experimental condition around the group mean, but with some idiosyncratic deviation from the group's JOL distribution as

much as from the group's calibration curve. For these two features to be added to the *JOLer*'s simulation tools, the *Simulation engine* form gives us slots to enter standard deviation scores (again as integer *x* such that *st.d.* = x/1000), one for the JOL distributions and another for the calibration curves. When these values are different from zero *JOLer*, before starting the simulation for a participant, computes an individual-specific distribution function and an individual-specific calibration curve for that participant. It draws on the per condition distribution and calibration, but randomly taking each JOL-associated value out from a pseudo-random normal distribution with mean the per-condition value and the standard deviation entered.

Finally, *JOLer* is presented as a graphical easy to use menu-driven application. The user can choose the language for its parameter-input forms and simulation-result reports (English/Spanish), and it provides extensive help and instructions to users. Additionally, the parameter values and the simulation results can be seen on the screen or they can be saved in an Excel 2003 file for storage and/or further inspections.

An example

As we said before, one of the most prominent phenomena in the metamemory research is the delayed-JOL effect (Nelson & Dunlosky, 1991). Table 1 shows the basic design of the simulation to analyze this experiment. For the sake of generality, we call here "control condition" to the immediate-JOL condition and "experimental condition" to the delayed-JOL condition. Typically, delaying the JOL increases the JOL resolution, but changing the JOL distribution too. The question here is whether the reported distribution alteration could explain the improved resolution. As each experimental condition is defined on two features in the simulation (i.e., JOL distribution and calibration curve), four simulation conditions are needed (2 x 2) to properly dissociate the contribution of each one of the features to the improved resolution.

Table 1. Basic simulation design: The numbers correspond to the condition identification number in JOLer. The letters between parentheses are those used by Weaver & Kelemen (1997, their Table 2) for an experiment in which the "control condition" stands for an immediate-JOL condition, and the "experimental condition" stands for a delayed-JOL condition.

	JOL distribution	
Calibration curve	Control	Experimental
Control	0 (A)	1 (B)
Experimental	2 (C)	3 (D)

Table 2 shows the JOL distribution and calibration curves reported by Weaver & Kelemen (1997) in a replication experiment of the delayed-JOL effect. We entered this scores in the *JOLer*'s *Simulation engine* form. Also we entered as global simulation parameters 4 conditions, 20 participants per condition, 60 trials per participant, 6 JOL bins, and 1000 experiment replications.

Table 2. Simulation 1: Parameter values used to replicate the Weaver and Kelemen (1997) simulation. The scores have been defined on the visual inspection of their Figures 1 and 2.

	JOL frequency (‰)		Correct recall (‰)	
JOL bin	Immediate-JOL	Delayed-JOL	Immediate-JOL	Delayed-JOL
0	80	400	240	20
20	150	210	310	200
40	220	35	530	530
60	270	35	600	780
80	150	80	670	840
100	130	240	750	920

JOLer yielded mean gammas .41, .58, .75, and .93 for condition 0, 1, 2, and 3, respectively. These correspond to their condition A, B, C, and D (see Weaver & Kelemen, 1997, Table 2). Certainly, our gamma means are clearly similar (when not the same) to those reported by Weaver & Kelemen (1997). It should also be noted that conditions A and D correspond to the experimental immediate and delayed JOL conditions, respectively: between them both the JOL distribution and the calibration curve are different. And they really exhibit the biggest between-condition gamma difference.

Particularly, interesting for the purpose of the simulation is the difference between condition A (control condition) and condition B (same calibration curve, but different distribution). Condition A, being the control condition, has been simulated under the JOL distribution and calibration curve empirically obtained by Weaver and Kelemen from their immediate-JOL data. Condition B has been simulated under the same calibration curve than condition A (i.e., no increase in calibration is assumed), but under a new JOL distribution, specifically, that empirically produced by their delayed-JOL condition. The interesting point here is that the mean gamma yielded by *JOLer* is clearly higher for simulated condition B. In other words, although the difference between condition A and D is considerably higher than that between A and B, the last one, being significant too, shows that the distributional alteration produced by delaying the JOLs does inflate the participant's gamma.

Checking *JOLer* with other examples

Weaver & Kelemen (2003) derived some predictions from the *Transfer Appropriate Monitoring* hypothesis. They wonder whether the similarity between JOL cues and the memory cues would improve JOL resolution as measured by the Goodman-Kruskal gamma correlation. Their experimental manipulations changed both the JOL distributions and the calibration curves. Interestingly, in addition to their experimental results they reported some simulations they run with the model implemented by *JOLer*. We could easily replicate their simulations as we did with the delayed-JOL effect conditions. For instance, let's assume that we want to replicate their simulation of the contrast between Conditions 1 and 4 for the recall group. From their Table 3 we can take the JOL frequency distributions for those conditions. Also

from the same table the corresponding calibration curves can be taken. Assume for the sake of argument that their Condition 1 is our *control condition* and their Condition 4 our *experimental condition*. Now, with the help of our Table 1 we can easily configure on *JOLer* the simulation replication for the 4 experimental conditions (2 JOL distributions by 2 calibration curves), with 60 participants per condition, 60 trials per participant, 6 JOL categories or bins, and 50 experiment replications as the global parameter values (see Weaver & Kelemen, 2003, p. 1063).

The results of this simulation -along with many others- reported by Weaver & Kelemen (2003) in their Table 4 can be compared to those yielded by *JOLer*, available in a supplementary file to this paper WK2003_12.xls. As a cross-checking for *JOLer* it can be seen how extremely close they are to each other. Other comparisons could be similarly done from the many simulation results reported in the paper by Weaver & Kelemen.

Conclusions

Whenever a researcher find out that in a metamemory experiment some experimental treatment has changed the JOL distribution along with JOL resolution, *JOLer* could help to easily disentangle the spurious contribution of the changed distribution to the increased (or decreased) resolution from the genuine effect on resolution.

Availability

Both the application as a zip file and an output sample named WK2003_14.xls are available for download from https://googledrive.com/host/0B_USuNXKEgWrcVdZamR5WXBPWnc.

Also *JOLer* is available as zip file upon request from the first author at no cost. The supplementary file WK2003_14.xls will also be provided. Before running the application, the content of the zip file should be unpacked in a *JOLer.jar* file and a *lib* folder; otherwise, it will not run properly. The *lib* folder included in the zip file has to remain as it is within the same folder as the executable *JOLer.jar*. *JOLer.jar* is a Java standalone application, so it is a cross-platform application, whenever the Oracle Java Virtual Machine version 7.0 or higher be installed on your computer (the Oracle's stable version today is 8.x). Once the zip file has been properly unpacked, the *JOLer.jar* executable file can be started as any other application in your system.

Acknowledgments.- The work described in this paper was done while the first author was partially supported by a grant from the Spanish Ministerio de Economía y Competitividad (PSI2013-47219-P).

Declaration of Conflicting Interests: The authors declared no potential conflicts of interests with respect to the research, author-ship, and/or publication of this article.

References

Arbuckle, T. Y. & Cuddy, L.L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psichology, 81*, 126-131.

Arnold, M.M., Higham, P.A., & Martín-Luengo, B. (2013). A little bias goes a long way: The effects of feedback on the strategic regulation of accuracy on formula-scored tests. *Journal of Experimental Psychology: Applied*, *19*, 383-402.

Begg, I., Duft, S., Lalonde, P. Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language, 28*, 610-632.

Benjamin, A.S. & Díaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R.A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73-94). New York, NY: Psychology Press.

Brown, R & McNeill, D. (1966). Te "tip of the tongue" phenomenon. *Journal of Verbal Learning & Behavior, 5*, 325-337.

Dunlosky, J. & Bjork, R.A. (Eds.) (2008). *Handbook of Metamemory and Memory*. Hove, NJ: Psychology Press.

Dunlosky, J. & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, Ca: Sage.

Dunlosky, J. & Nelson, T.O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 373-380.

Dunlosky, J. & Nelson, T.O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effect of various activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545-565.

Finn, B. & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *33*, 238-244.

Finn, B. & Metcalfe, J. (2014). Overconfidence in children's multitrial judgments of learning. *Learning and Instruction*, 32, 1-9.

Glenberg, A.M. & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15,* 84–93.

Glucksberg, S. & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 311-325.

González, R. & Nelson, T.O. (1996). Measuring ordinal association in situations that contain tied scores. Psychological Bulletin, *119*, 159-165.

Goodman, L.A. & Kruskal, W.H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732-764.

Hart, J.T. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psichology, 56,* 208-216.

Hart, J.T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6,* 685-691.

Kelley, C. M. & Jacoby, L. L. (1996). Memory attributions: Remembering, knowing, and feeling of knowing. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 287-307). Mahwah, NJ: Erlbaum.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609-639.

Koriat, A. (1997). Monitoring one's own knowledge during study: a cue utilization approach to judgments of learning. *Journal of Experimental Psichology: General, 126,* 349-370.

Koriat, A. & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517.

Koriat, A. & Levy-Sadot, R. (2001). The Combined Contributions of the Cue-Familiarity and Accessibility Heuristics to Feelings of Knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 34-53.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 595-608.

Koriat, A. & Nussinson, R. (2009). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 338–1343.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgment of learning exhibit increased underconfidence-with-practice. *Journal of Experimental Psychology: General*, *131*, 147-162.

Liberman, V. & Tversky, A. (1993). On the evaluation of probability judgments: Calibration, discrimination, and monotonicity. *Psychological Bulletin*, *114*, 162-173.

Lovelace, E. A. (1984). Metamemory: Monitoring Future Recallability During Study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756-766.

Luna, K., Higham, P.A., & Martín-Luengo, B. (2011). Regulation of memory accuracy with multiple answers: The plurality option. *Journal of Experimental Psychology: Applied*, *17*, 148-158.

Masson, M.E.J. & Rotello, C.M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of associations: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 509-527.

Metcalfe, J. & Finn, B. (2008). Familiarity and Retrieval Processes in Delayed Judgments of Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1084–1097.

Metcalfe., J., Schwartz, B. L. & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851-861.

Muruyama, K., Sakaki, M., Yan, V.X., & Smith, G.M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1287-1306.

Nelson, T.O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109-133.

Nelson, T.O. y Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting recall: The "delayed-JOL effect". *Psychological Science*, *2*, 267-270.

Nelson. T.O., Leonesio, R.J ., Landwehr. R.S., & Narens. L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 279-287.

Nelson, T. O. & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and felling-ofknowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19,* 338–368.

Nelson, T.O., y Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26,* 125-173.

Pyc, M.A., Rawson, K.A., & Aschenbrenner, A.J. (2014). Metacognitive monitoring during criterion learning: When and why are judgments accurate. *Memory & Cognition, 42,* 886–897.

Reder, L.M. (Eds.) (1996). *Implicit Memory and Metacognition*. Mahwah, NJ: Lawrence Erlbaum.

Rhodes, M. G. & Tauber, S.K. (2011). The Influence of Delaying Judgments of Learning on Metacognitive Accuracy: A Meta-Analytic Review. *Psychological Bulletin, 137*, 131–148.

Ruiz, M. (2004). *Las caras de la memoria.* Madrid: Pearson-Prentice Hall.

Serra, M.J. & Ariel, R. (2014). People use the memory for past-test heuristic as an explicit cue for judgments of learning. *Memory & Cognition, 42,* 1260–1272.

Shanks, L.I. & Serra, M.J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin & Review*, *21*, 445-453.

Sikström, S. & Jönsson, F. U. (2005). A model for stochastic drift in memory strength to account for judgments of learning. *Psychological Review. 112(4),* 932-950.

Son, L. K. & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition, 33*, 1116-1129.

Sundqvist, M.L., Todorov, I., Kubik, V., & Jönsson, F. U. (2012). Study for now, but judge for later: Delayed judgments of learning promote long-term retention. *Scandinavian Journal of Psychology, 53,* 450–454.

Underwood, G. (Coords.) (1966). *Implicit cognition.* Oxford, U.K.: Oxford University Press.

Vesonder, G. T. & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory & Language, 24,* 363-376.

Weaver, C.A. & Kelemen, W.L. (1997). Judgmens of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science*, *8*, 318-321.

Weaver, C.A. & Kelemen, W.L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1058-1065.

Weaver, C. A., Terrell, J. T., Krug, K. S., & Kelemen, W. L. (2008). The Delayed JOL Effect with very long delays: Evidence from flashbulb memories. In J. Dunlosky and R. A. Bjork (Eds.), *A handbook of memory and metacognition* (pp. 155-172). Hillsdale, NJ: Lawrence Erlbaum Associates.

(Article received: 28-03-2015; revised: 02-10-2015; accepted: 19-10-2015)