The thousand-question Spanish general knowledge database

Francisco Buades-Sitjar[1], Roger Boada[2], Marc Guasch[2], Pilar Ferré[2], José Antonio Hinojosa[3,4,1], Marc Brysbaert [5], and Jon Andoni Duñabeitia[1,6*]

[1] Centro de Ciencia Cognitiva - C3, Universidad Antonio de Nebrija, Madrid, Spain

[2] Universitat Rovira i Virgili, Department of Psychology and CRAMC, Tarragona, Spain.

[3] Instituto Pluridisciplinar, Universidad Complutense de Madrid, Madrid, Spain.

[4] Dpto. Psicología Experimental, Procesos Cognitivos y Logopedia, Universidad Complutense de Madrid, Madrid, Spain

[5] Ghent University, Ghent, Belgium

[6] Department of Languages and Culture, The Arctic University of Norway, Tromsø, Norway

*Correspondence:

Jon Andoni Duñabeitia

jdunabeitia@nebrija.es

Centro de Ciencia Cognitiva – C3

Universidad Antonio de Nebrija

Santa Cruz de Marcenado 27

Madrid, 28015, Spain

Abstract

General knowledge questionnaires have been ubiquitously used to study a wide variety of phenomena, such as illusory truth, error correction and tip-of-the-tongue situations. However, their normings are highly restricted to the territory and the time period they in which they were obtained. This requires that new normings are obtained for each new territory in which they be used. Here, we present a new set of 1364 general knowledge questions normed for a Spanish population. The questions span a total of 37 different fields of knowledge and an extensive range of difficulty levels. They are formulated in a multiple-choice format, and pick rates for the correct answer as well as for the three incorrect response options are provided. We hope that a database of such size and flexibility will prove to be a useful research tool for the Spanish community.

Keywords: general knowledge, Spanish norms, English translation, metacognition, illusory truth

Introduction

Our relationship to general knowledge —information about facts of a varied nature that are considered as relatively widely-known by a population— has changed over the last few decades. For instance, the labor market currently favors specialization, with most jobs requiring some kind of post high school degree, which greatly reduces the need of having a wide collection of facts stored in one's memory. Furthermore, the widespread use of the Internet has also had a paradoxical influence on our views towards this type of knowledge, as even though most pieces of general information are now easily accessible to everyone at any time, this availability also makes it unnecessary to retain them in our memory (see the review on distributed cognition by Michaelian and Sutton, 2013, for a deeper dive into this topic). Yet, in spite of all this, general knowledge is still highly valued by individuals, and is usually regarded as one of the hallmarks of what we consider to be a "cultured person". Thus, it comes as no surprise that psychologists have remained interested in the acquisition and retainment of this type of data over the years (e.g., Bäckman and Lipinska, 1993; Murayama and Kuhbandner, 2011; Coane and Umanath, 2019) and that much effort has been invested into creating and updating scales as a means to measure a person's general knowledge (Duñabeitia et al., 2016; Nelson and Narens, 1980; Jalbert, Newman and Schwarz, 2019; Tauber, Dunlosky, Rawson, Rhodes and Sitzman, 2013; Martín-Luengo, Zinchenko, Alekseeva and Shtyrov, 2020).

To date, the most widely used set of general knowledge datasets in psychological research has been the one created by Nelson and Narens (1980). It is comprised of 300 general knowledge questions, spanning a wide variety of topics and levels of difficulty, which were answered by 270 students from North American universities. The set has been used to study a myriad of different psychological phenomena, such as illusory truth (Fazio, Brashier, Pain and Marsh, 2015), error correction (Sitzman, Rhodes and Tauber, 2014; Sitzman, Rhodes, Tauber and Liceralde, 2015), self-performance evaluation (Jackson and Greene, 2014; Weinstein and

Roedinger, 2010) and tip-of-tongue situations (Schwartz, 2010), cementing it as the largest and most prevalent database of this kind.

Nevertheless, in spite of its widespread use, Nelson and Narens' (1980) database, as well as any other set of general knowledge questions in existence, face two main issues. The first one is their applicability across time. As years pass, society changes, new technologies are developed and new culturally relevant events take place. Hence, what is considered as general knowledge at a certain point in time might not be at another. Despite Nelson and Narens' (1980) best efforts to select a list of ageless questions for their database, Tauber et al. (2013) noted that the passing of time had rendered the norming of several of the questions invalid, as either the proportion of people who knew the correct answer had shifted, or the correct answer itself had changed over the three-decade gap. Thus, they set to update Nelson and Narens' (1980) database in order to reflect the current state of the questions. To this end they asked 671 students from two North American universities to answer the same 300 questions, and they also added a new set of data for each of the items. This included self- and peer-confidence judgements (i.e., how likely the person thought they or their peers were of knowing the correct answer on a scale from 0 to 100) and, most importantly, commission errors (i.e., the most common incorrect answers for each of the questions). Although to a much lesser extent than Nelson and Naren's (1980), the Tauber et al. (2013) norms have been used in a variety of studies as well (e.g., Bashier, Umanath, Cabeza and Marsh, 2017; Chua and Bliss-Moreau, 2016; Coane and Umanath, 2019), and it is generally considered a more age-appropriate alternative to the original.

The second limitation is their cross-regional applicability. The cultural differences of each territory result in differences in what is considered as general knowledge as well. For instance, while the answer to the question "What is the capital of Delaware?" might be considered as widespread, ageless knowledge in the US, it is probably seen as an obscure piece

of information in many other countries. This means that the norms for each of the questions in the set created by Nelson and Narens (1980) in the US —as well as in the Tauber et al. (2013) update— do not necessarily reflect the reality of general knowledge in other countries. Thus, a simple translation of the questions is not enough to apply them to a foreign population, and a new norming process is required. In this line, Duñabeitia et al. (2016) created a Spanish adaptation of the Tauber et al. (2013) set by eliminating all questions that were not considered to have cross-cultural applicability, and they collected norms with a Spanish sample. However, while this made the database applicable to a Spanish population, it also meant the set was reduced from 300 questions to only 132. As pointed out by Jalbert et al. (2019), only having access to such a limited number of questions poses a threat to the viability of the database if used repeatedly, as it increases the likelihood that a given person has already been exposed to the same question, hampering the validity of any results that might be found.

To address these issues, we present a new set of 1364 general knowledge questions that have been selected to cover a wide variety of seemingly country-unspecific topics, and that have been normed using a Spanish population. Following Nelson and Narens (1980), we chose an array of questions from many different topics and difficulty levels to enable the validity of the norming to last for as long as possible, and in this Data Report we provide the ratios for each correct answer and commission error.

Our database has two key differences relative to Nelson and Narens (1980). The first one is that it contains a considerably larger number of questions. This helps ensure that participants will not be likely to have been previously exposed to the same questions, even after taking part in repeated experiments using the same database. The second one is that we strayed from Nelson and Narens' (1980) format of one-word, open-ended answers in favor of a four-option, multiple choice format. This allows for the automatic correction of the questions, as it does not require a human check for spelling mistakes, greatly easing the workload. It also

allows us to include multiple-word answers, widening the type of questions that can be included. Moreover, as noted by Jalbert et al. (2019), several studies examining statement recognition and perceived truth have adapted the Nelson and Narens (1980) questions to form true ("Photosynthesis is the name of the process by which plants make their food") and false ("Chemosynthesis is the name by which plants make their food") statements (e.g., Brashier at al., 2017; Fazio et al., 2015; Newman, Garry, Bernstein, Kantner and Lindsay, 2012). The incorrect answers used as part of the false statements do not necessarily follow any kind of norming between different studies, hampering the cross-validity of any possible findings. Other studies such as Chua and Bliss-Moreau (2016), Coane and Umanath (2019) and Schwarz (2010) have also changed the questions from the two databases from an open-ended into a multiple-choice format for their own purposes; again, with no regard for cross-study consistency. Hence, by providing a set of incorrect answers ourselves —along with their respective prevalence— we can aid in solving this problem. We hope that these changes will help us provide the scientific community with a durable general knowledge measure to use in research. In the following section, we describe the item and participant selection process, as well as provide details of the norming process.

Method

Participants

The original sample comprised 400 non-migrant Spanish participants —all of whom spoke Spanish as their native language— who voluntarily took part in this study. They were offered the possibility of winning one 25€ Amazon coupon per every 25 participants. Considering that the data collection was completed online, and in order to grant high data quality, a series of a priori selected filters were applied. On one hand, participants who took an average of less than four seconds to answer each question were removed from the analysis.

Similarly, participants who took longer than three standard deviations from the average to complete the task were also discarded. This brought the final sample down to 385 participants (95 males, 288 females and two people who chose not to disclose their gender; mean age = 22.24, SD = 8.24; mean socioeconomic status measured on a 1-to-10 scale = 5.88, SD = 1.31). Participants were all college students from a wide variety of majors, recruited from three different Spanish universities (Universidad Complutense de Madrid and Universidad Nebrija in Madrid, and Universitat Rovira i Virgili in Catalonia) in order to ensure the cross-territorial validity of the norming. Before taking part in the data collection, participants provided informed consent according to the Spanish regulations. The protocol was validated by the Ethics Board of the Universidad Nebrija (approval code UNNE-2020-008).

Materials

A total of 1364 general knowledge questions and their answers were extracted from a wide array of different sources, ranging from academic databases such as Tauber et al. (2013) to question-based online games such as *funtrivia.com* and *triviachamp.com*. Their content spanned a total of 37 different fields of knowledge (e.g, architecture, biology, history, philosophy and sports), as rated by three independent judges. For each question, four possible answers were created (the correct answer and three credible alternatives). The incorrect but plausible alternative response options acted as lures when inserted in the multiple-choice format. The veracity of all question-answer pairs was then manually reviewed, and the Wikipedia articles that could confirm the answer were also found.

Procedure

Data collection was obtained remotely through the use of *Gorilla* software (Anwyl-Irvine, Massonnié, Flitton, Kirkham and Evershed, 2020). Before presenting the questions, participants indicated their gender, age, autonomous community of origin and their

socioeconomic status (SES), which was obtained through the use of the MacArthur Scale of Subjective Social Status (Adler and Stewart, 2007). Once participants provided their demographic information, they proceeded with the task. The items were divided into four lists, with each participant completing one of them. All items were presented in a random order, and each question was presented together with the four possible answer options. Participants were told to choose the one they believed to be correct by clicking on it. The task lasted for approximately an hour, and several breaks were weaved in every few trials to prevent fatigue.

Results

Both the Spanish version and the English translations of the questions are provided in https://figshare.com/articles/dataset/The_thousand-question_Spanish_general_knowledge_database/13041803 . The dataset contains a number from 1 to 1364 that corresponds to each of the questions (*QUESTION NUMBER)*, followed by the question itself (*QUESTION TEXT)*. Next, the correct answer (*CORRECT ANSWER)* and the other three incorrect response options (*LURE 1, LURE 2* and *LURE 3)* are presented, followed by the average pick rate for each of them (*MEAN CORRECT* and *MEAN LURE 1,2,3)* ordered in decreasing order (namely, the most commonly picked incorrect answer to each question is provided first, followed by the second most commonly selected erroneous answer, and followed by the least selected one. Finally, the field of knowledge of the question (*CATEGORY)* and the link to the Wikipedia article in which to find the correct answer (*LINK)* are also provided in the last two columns.

A closer look at the distribution of the mean accuracy rates for each question (mean = 50%, SD = 23.7%) shows that the database as a whole has a healthy distribution of questions in all ranges of difficulty (see Figure 1). Table 1 displays a per-category breakdown of the accuracy rates, the highest being that of *Biology* (67.96%) and the lowest being that of

Inventions (35.02%). As seen, the range of the mean accuracy rates per category is relatively narrow, suggesting certain homogeneity in the degree of difficulty across topics. All of this, combined with the great number of items included, ensures that researchers using this database will very likely be able to construct an appropriate subset of questions regardless of the specific difficulty needs of their study. Furthermore, we examined the split-half reliability of the test questions by computing the Spearman-Brown coefficient between the responses given by the even and the odd participants, and this resulted in a 0.89 coefficient. Such a high score ensures a high degree of consistency in the responses to each of the questions.

- Insert Table 1 and Figure 1 around here -

Discussion

The present Data Report introduces a new database of general knowledge questions akin to that of Nelson and Narens (1980) and its updates and adaptations by Tauber et al. (2013) or Duñabeitia et al. (2016). In a similar fashion to its predecessors, the items included in this database span a wide variety of topics, ensuring that no advantage is conferred to people who are well-versed in a specific field of knowledge. Likewise, the sample used in the norming process was extracted from three different universities. This endorses the cross-territorial validity of the dataset.

The current Data Report also addresses some of the issues with previous databases of the same kind. First, it comprises a much greater number of items than its predecessors. Second, the multiple-choice format allows for a much quicker distribution of the questions, as no manual check for answer spelling mistakes is required. Additionally, the manner in which the questions are structured (*question particle* + *question content*), significantly facilitates turning the questions into standardized true and false statements. This will help ensure greater

consistency in studies examining phenomena such as error correction and illusory truth, which heavily rely on these kinds of statements.

Our database, however, is not without its limitations. On the one hand, while we provide the English translation of all items in the set, it is critical to keep in mind that new norms for these items are highly recommended if they are to be used by researchers studying non-Spanish samples, since their results could likely change from country to country. On the other hand, despite our attempts at selecting a set of ageless questions, no database of this kind is completely impervious to the passing of time. Hence, while the norming in this Data Report faithfully represents the state of the questions at the moment of publishing, the same might not be true as years pass. Nevertheless, the division of the items into the different fields of knowledge to which they belong greatly helps reduce the impact of time on the database: while the norming of questions belonging categories such as *Music* or *Television* is more vulnerable to the passing of time, the norming of questions in categories such as *Biology, Physics* and *Math* is very likely to remain the same for many years. This means that, by selecting the appropriate categories, it is possible to bypass the impact that the passing on time might have on the norming.

Finally, it should be kept in mind that the test sample of the current study is exclusively composed of young adults, all being university students, thus creating an inherent bias in the results and representing a relatively homogeneous population. In fact, when running a linear regression analysis on the mean accuracy scores on the basis of SES —the only demographic variable showing certain level of variability in the sample—, we found it had no predictive power ($t<1$ and $p>0.45$). While previous studies have found SES to be a powerful predictor of IQ scores that include general knowledge measures (e.g., Schoon, Jones, Cheng and Maughan, 2012; Von Stumm and Plomin, 2015), the relationship between SES and intelligence seems to be mediated by educational level (see Altschul, 2012; Ritchie and Tucker-Drob, 2018).

Consequently, finding a limited impact of SES in the current study is not entirely surprising, given that our sample is comprised entirely by university students. Besides, it is worth noting that our test sample included three times more females than males. Hence, in light of these issues, caution is advised to extrapolate these scores to the general population, and a larger scale study with a more representative sampling of participants is recommended to allow for exploring the influence of sociodemographic factors on general knowledge. In spite of these issues, we are confident that the current dataset will provide researchers with a flexible and reliable tool to assess general knowledge, as well as studying other question- and statement-related phenomena.

References

Adler, N., & Stewart, J. (2007). The MacArthur scale of subjective social status. MacArthur Research Network on SES & Health. Retrieved from http://www. macses. ucsf. edu/Research/Psychosocial/subjective. php.

Altschul, I. (2012). Linking socioeconomic status to the academic achievement of Mexican American youth through parent involvement in education. *Journal of the Society for Social Work and Research*, *3*(1), 13-30.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods, 52*(1), 388-407.

Bäckman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia*, *31*(4), 335-345.

Brashier, N. M., Umanath, S., Cabeza, R., & Marsh, E. J. (2017). Competing cues: Older adults rely on knowledge in the face of fluency. *Psychology and aging*, *32*(4), 331.

Chua, E. F., & Bliss-Moreau, E. (2016). Knowing your heart and your mind: The relationships between metamemory and interoception. *Consciousness and cognition*, *45*, 146-158.

Coane, J. H., & Umanath, S. (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. *Journal of Memory and Language*, *107*, 152-168.

Duñabeitia, J. A., Griffin, K. L., Martín, J. L., Oliva, M., Sámano, M. L., & Ivaz, L. (2016). The Spanish General Knowledge Norms. *Frontiers in psychology*, *7*, 1888.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.

Jackson, A., & Greene, R. L. (2014). Impression formation of tests: Retrospective judgments of performance are higher when easier questions come first. *Memory & cognition, 42*(8), 1325-1332.

Jalbert, M., Newman, E., & Schwarz, N. (2019). Trivia claim norming: Methods report and data. ResearchGate.

Martín-Luengo, B., Zinchenko, O., Alekseeva, M., Y Shtyrov, Y. (2020). Russian norms for 500 general-knowledge questions. Front. Psychol. 11:02269. doi: 10.3389/fpsyg.2020.545304

Michaelian, K., & Sutton, J. (2013). Distributed cognition and memory research: History and current directions. *Review of philosophy and psychology*, *4*(1), 1-24.

Murayama, K., & Kuhbandner, C. (2011). Money enhances memory consolidation–But only for boring material. *Cognition*, *119*(1), 120-124.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of verbal learning and verbal behavior*, *19*(3), 338-368.

Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review*, *19*(5), 969-974.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychological Science*, *29*(8), 1358–1369.

Schoon, I., Jones, E., Cheng, H., & Maughan, B. (2012). Family hardship, family instability, and cognitive development. *J Epidemiol Community Health*, *66*(8), 716-722.

Schwartz, B. L. (2010). The effects of emotion on tip-of-the-tongue states. *Psychonomic Bulletin & Review*, *17*(1), 82-87.

Sitzman, D. M., Rhodes, M. G., & Tauber, S. K. (2014). Prior knowledge is more predictive of error correction than subjective confidence. *Memory & Cognition*, *42*(1), 84-96.

Sitzman, D. M., Rhodes, M. G., Tauber, S. K., & Liceralde, V. R. T. (2015). The role of prior knowledge in error correction for younger and older adults. *Aging, Neuropsychology, and Cognition*, *22*(4), 502-516.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior research methods, 45(4)*, 1115-1143.

Von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence*, *48*, 30-36.

Weinstein, Y., & Roediger, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition, 38*(3), 366-376.

ACKNOWLEDGEMENTS

This research has been partially funded by grants PID2019-107206GB-I00, PGC2018-098558-B-I00, PGC2018-097145-B-I00 and RED2018-102615-T from the Spanish Government and H2019/HUM-5705 from the Comunidad de Madrid. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors thank Lela Ivaz for her contribution to the initial stages of this project.

Figure 1. Histograms and density of the pick rates for each of the four possible response options. The lures are organized according to the mean pick rate in decreasing order (with Lure 1 being the most selected incorrect answer).

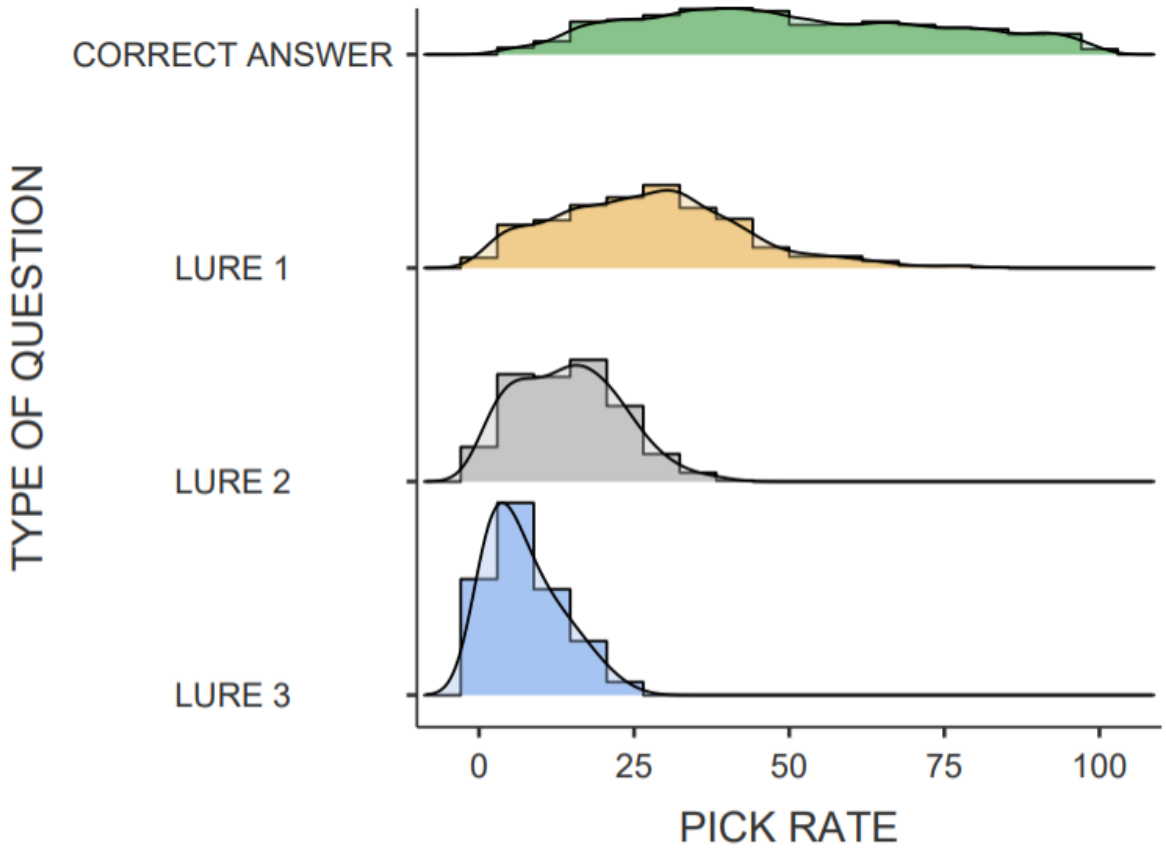

Table 1

Percent Mean Accuracy and Standard Deviation per Category (Ordered From Highest to Lowest)

Category	Mean Accuracy	SD	Category	Mean Accuracy	SD
Biology	67.96	18.78	Movies	50.5	21.34
Human Body	65	25.87	Music	49.14	22.84
Psychology	63.21	24.47	Technology	48.67	24.39
Math	60.74	26.56	Chemistry	48.46	24.79
Food	58.78	25.27	Religion	48.43	23.56
Achitecture	58.26	22.46	Phylosophy	47.96	21.96
Medicine	57.49	23.41	Mythology	47.45	19.85
Art	56.87	21.52	Literature	47.14	20.14
Economy	56.81	24.88	Transportation	45.52	23.37
Measurements	55.97	26.53	Organizations	44.97	25.68
Records	52.14	26.54	Physics	43.98	21.08
Botany	51.92	23.11	World	42.43	16.73
Astronomy	51.64	25.53	Classical music	41.81	21.44
Television	51.55	20.45	Astronauts	40.35	23.2
Europe	51.31	22.71	Brands	40.07	23.95
Sports	51.13	22.57	Linguistics	39.11	19.2
Politics	50.88	22.92	History	37.3	20.91
Zoology	50.72	25.33	Inventions	35.02	16.89
Geography	50.55	24.68			