White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/153185/

Version: Accepted Version

Article:
Ferrero, M, Hardwicke, TE, Konstantinidis, E et al. (2020) The effectiveness of refutation texts to correct misconceptions among educators. Journal of Experimental Psychology: Applied, 26 (3). pp. 411-421. ISSN: 1076-898X

https://doi.org/10.1037/xap0000258

©American Psychological Association, 2020. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: https://doi.org/10.1037/xap0000258

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Running Head: REFUTATION TEXTS IN EDUCATION

IN PRESS

Journal of Experimental Psychology: Applied

21/10/2019

The effectiveness of refutation texts to correct misconceptions among educators

Marta Ferrero[1], Tom E. Hardwicke[2,3], Emmanouil Konstantinidis[4,5], & Miguel A. Vadillo[6,7]

[1]*Facultad de Educación, Universidad Complutense de Madrid, Spain*

[2]*Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Charité – Universitätsmedizin Berlin, Germany*

[3]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, USA*

[4]*Centre for Decision Research, University of Leeds, UK*

[5]*Department of Psychology, University of Warwick, UK*

[6]*Primary Care and Public Health Sciences, King's College London, UK*

[7]*Departamento de Psicología Básica, Universidad Autónoma de Madrid, Spain*

Mailing address:

Marta Ferrero

Facultad de Educación, Centro de Formación del Profesorado

Universidad Complutense de Madrid

c/ Rector Royo Villanova s/n, Ciudad Universitaria, 28040, Madrid, Spain

e-mail: maferr16@ucm.es

Public Significance Statement

Existing evidence indicates that teachers endorse a considerable number of misconceptions about education and neuroscience that hinder the adoption of evidence-based practices in the classroom. The present study found that refutation texts can be an effective means to correct educational misconceptions, although their effects were only temporary and did not change teachers' intention to adopt educational practices based on those erroneous ideas.

Abstract

Teachers around the world hold a considerable number of misconceptions about education. Consequently, schools can become epicenters for dubious practices that might jeopardize the quality of teaching and negatively influence students' wellbeing. The main objective of this study was to assess the efficacy of refutation texts in the correction of erroneous ideas among in-service teachers. The results of Experiment 1 indicate that refutation texts can be an effective means to correct false ideas among educators, even for strongly endorsed misconceptions. However, the results of Experiment 2 suggest that these effects may be short-lived. Furthermore, attempts to correct misconceptions seemed to have no beneficial effect on teachers' intention to implement educational practices that are based on those erroneous beliefs. The implications of these results for the training of pre-service and in-service teachers are discussed.

Keywords: Teachers; Misconceptions; Refutation texts; Origin of information; Strength of belief

The prevalence of misconceptions about education among school teachers and college teachers is high. For instance, the belief that adapting teaching to the "learning styles" of students improves learning is shared by more than 85% of teachers in different countries, such as The Netherlands, United Kingdom, Greece, Argentina, Chile, Peru, Turkey, and Spain. Similarly, more than 67% of educators worldwide believe that differences in hemispheric dominance (left brain, right brain) can help explain individual differences amongst learners (Dekker, Lee, Howard-Jones, & Jolles, 2012; Deligiannidi & Howard-Jones, 2015; Ferrero, Garaizar, & Vadillo, 2016; Gleichgerrcht, Luttges, Salvarezza, & Campos, 2015; Hunter & Lloyd, 2018; Karakus, Howard-Jones, & Jay, 2015; Morehead, Rhodes, & DeLozier, 2015). However, both statements have been broadly discredited by research (Coffield, Moseley, Hall, & Ecclestone, 2004; Corballis, 1999; Hines, 1987; Pashler, McDaniel, Rohrer, & Bjork, 2008). The situation is not substantially different for in training teachers (e.g. Fuentes & Risso, 2015).

The dissemination of these misconceptions among teachers through workshops, conferences, and educational materials promotes the use of pseudoscientific practices in the classroom (Busso & Pollack, 2014; Goswami, 2006) and might impede the adoption of evidence-based practices, compromising the quality of education. For example, a sizeable number of schools employ methods which have been shown to be ineffective by the scientific community, such as Brain-Gym® (Hyatt, 2007) or "learning styles" (Pashler et al., 2008). These interventions usually require modification of curricular programs, time in training courses for educators, adaptation of teaching materials and learning spaces, and a considerable amount of money, typically from limited public sources (Goldacre, 2006). Consequently, they divert investment from well-grounded methodologies. Even worse, the implementation of these dubious practices can translate into adverse consequences for the students. To mention just a few examples, the popular "patterning therapy", a series of

exercises aimed at improving the neurological organization of impaired children, not only wastes valuable time, it also seriously compromises children's self-confidence, sibling relationships, and parents' financial resources (American Academy of Pediatrics, 1999; Vergara, Martínez, Martínez-Sahuquillo, & Echevarría, 2011). Similarly, the widely used "auditory integration training", which focuses on a broad range of disorders with special emphasis on autism (American Academy of Pediatrics, 1998), involves a high monetary cost, and may cause undesirable side effects in children (American Speech-Language-Hearing Association, n.d.; New York State Department of Health, 1999).

The impact of misinformation in our society is not restricted to educational issues. For example, many people believe that global warming is not due to CO_2 emissions generated by human activity (Lewandowsky, Oreskes, Risbey, Newell, & Smithson, 2015) or that vaccines may cause severe adverse side effects for children (Freed, Clark, Butchart, Singer, & Davis, 2009). Faced with this reality, several studies have been undertaken to determine the best way to correct misconceptions. Most of this research has been conducted in the domains of politics (i.e., Nyhan & Reifler, 2010; Nyhan, Reifler, & Ubel, 2013), health (i.e., Nyhan & Reifler, 2015; Nyhan, Reifler, Richey, & Freed, 2014), and school and college education (i.e., Guzzetti, Williams, Skeels, & Wu, 1997; Hynd, McWorther, Phares, & Suttles, 1994; Kowalski & Taylor, 2009; Maria & MacGinitie, 1987; Tippett, 2010). Taken collectively, these studies suggest that, once accepted, misconceptions tend to be quite resistant to change and that not all correction methods are effective (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Some evidence suggests that attempts to correct misinformation can actually backfire; that is, instead of mitigating erroneous ideas, they can inadvertently strengthen them (e.g., Nyhan & Reifler, 2010, 2015; Nyhan et al., 2013). However, several studies have been unable to replicate this finding (Haglin, 2017; Swire, Ecker, & Lewandowsky, 2017; Wood & Porter, 2019).

One of the reasons why misconceptions are so difficult to correct is that it usually requires modification of a complex network of self-reinforcing beliefs or 'worldview' (Lewandowsky et al., 2012). Not surprisingly, individuals are often reluctant to abandon previous conceptions in spite of their being inaccurate (Chi, 2005) and they persist in relying on misinformation even when they can recall a correction, a phenomenon known as the *continued influence effect* (Johnson & Seifert, 1994).

In this context, refutation texts are emerging as a promising means to correct misinformation. Refutation (or refutational) texts are defined as those that describe a common theory, belief, or idea, refute it, and offer an alternative that is shown to be more satisfactory (Guzzetti, 2000). Presumably, highlighting the inconsistencies between false beliefs and correct information might prepare the ground for a conceptual change (Hynd, 2001; Kendeou & van den Broek, 2007). A second merit of refutation texts is that they provide an alternative to replace the original misinformation. This element is critical to allow people to update their knowledge and prevent them from relying on the same misconception in the future (Johnson & Seifert, 1994; Ecker, Lewandowsky, & Tang, 2010; Wilkes & Leatherbarrow, 1988). Ideally, the alternative explanation should be plausible, that is, it must help to resolve the problem generated and also be consistent with other knowledge (Posner et al., 1982). In addition, it should never be more complicated to understand than the original idea it attempts to replace (Lombrozo, 2007). Finally, refutation texts seem to be more effective if they include detailed information (Tippet, 2010) and if they make the reader suspicious about the reliability of the original misconception or its source (Lewandowsky, Stritzke, Oberauer, & Morales, 2005). Consistent with this, people seem to be more willing to revise their erroneous beliefs when it is explained that these ideas do not originate from experts or trustworthy sources (Guillory & Geraci, 2013). Based on this evidence, we can expect that a refutation text will be more effective if it not only addresses the reason why a piece of information is

false but also highlights that the original statement of the misconception was not based on solid evidence or sound reasoning to begin with.

According to some studies, the effectiveness of refutation texts and other debunking methods varies according to the extent to which misinformation is endorsed. Some research has suggested that firmly held beliefs are apparently the most resistant to change (Lewandowsky et al., 2012; Nyhan & Reifler, 2015), to such an extent that some authors have recommended to target only people with moderate rather that strong beliefs when it comes to refuting incorrect information (Ecker, Swire, & Lewandowsky, 2014). This finding is consistent with a long tradition of research on motivated reasoning showing that the interpretation of new information is usually biased by the pre-existing beliefs and attitudes of the recipient (Kunda, 1990). In contrast, other studies have shown that high-confidence errors are easier to correct than low-confidence ones (Butterfield & Metcalfe, 2001; see also Ecker, Lewandowsky, Fenton & Martin, 2014) and that refutation texts might be especially beneficial for the correction of the former (van Loon, Dunlosky, van Gog, van Merriënboer, & de Bruin, 2015). Many misconceptions are deeply integrated into the daily practice of teachers but, to the best of our knowledge, the relevance of this point in the context of educational misinformation among in-service educators has not been addressed by previous research.

In sum, in-service teachers endorse a considerable number of educational misconceptions, with potentially serious consequences. Despite the importance of this topic, only a modest number of studies have been conducted to address teachers' educational misconceptions and all of them have been exclusively focused on pre-training teachers (e.g. Gill, Ashson, & Algina, 2004; Salisbury-Glennon & Stevens, 1999). Unlike them, in-service teachers may have been exposed to a greater or lesser extent to different misconceptions and, consequently, the familiarity and confidence they have towards these misconceptions may

influence the effectiveness of correction efforts. The main objective of the present study was to test the effectiveness of refutation texts in dispelling widespread myths about education among in-service teachers. We also examined whether the effectiveness of these texts was influenced by the inclusion of information discrediting the origin of the erroneous ideas.

Experiment 1

In this experiment, we measured the extent to which participants endorsed a number of misconceptions before and after exposing them to different refutation texts. Specifically, we examined the effectiveness of these texts depending on the inclusion or not of information discrediting the origin of the misconception, as well as the pre-existing endorsement of the misconceptions addressed in them. Our hypotheses were that (a) exposing teachers to refutation texts would be more effective in reducing or eliminating misconceptions than doing nothing, and that (b) refutation texts would be even more effective when they explicitly discredited the origin of misinformation. Additionally, since the effectiveness of refutation texts might depend on the degree of familiarity or endorsement of incorrect information, we were also interested in testing the impact of these variables on the effectiveness of the refutation texts. In this vein, we hypothesized that (c) teachers with more years of experience in schools, and hence probably more exposed to misconceptions, would be more resistant to corrections, and that (d) the most strongly endorsed misconceptions among educators would be more resistant to change.

Method

Participants

Because this is a difficult sample to recruit, we aimed to test as many participants as possible instead of basing sample size on a power calculation. The sample included 45

teachers (40 female) from 12 different schools of Spain. The mean age of participants was 39.08 (*SD* = 10.12). Participants were special education teachers (24.44%), kindergarten teachers (42.22%), primary school teachers (13.33%), secondary school teachers (4.44%), vocational education teachers (4.44%), and teachers who worked in more than one level of education (11.11%). The sample was recruited from public (35.55%), private (11.11%) and state schools (53.33%). The average teaching experience of the participants was 13.4 years (*SD* = 10.12).

To participate in the experiment, teachers were contacted by email invitation to schools randomly selected from the official list of schools of the Spanish Ministry of Education (https://www.educacion.gob.es/centros/home.do). They were told that the aim of the study was to determine the best way to transmit knowledge about neuroscience and education to in-service teachers. After agreeing to participate in the study, schools received an email with the instructions to complete the study. One day before the start date of each phase, headmasters received a reminder they should forward to the teachers of their schools. In this reminder (available at https://osf.io/umz5w/) an explanation about the nature and duration of the tasks was provided. All participants in Experiments 1 and 2 provided informed consent and the studies were approved by the KCL ethics committee (ref. LRS-15/16-2833).

Design and Procedure

To maximise power, we used a fully within-participants design. The experiment consisted of two phases. During Phase 1, participants completed an on-line questionnaire assessing the extent to which they endorsed a number of statements related to education and neuroscience applied to education (for more details, see the Materials section). This level of endorsement, measured by means of a Likert scale (see below), served as our main dependent variable in Phases 1 and 2. Eighteen of the items included in this questionnaire referred to

common misconceptions (e.g., “We only use 10% of our brain”) and the other 18 were correct statements about education (e.g., “Boys have bigger brains than girls.”). Of particular importance, nine of the 18 misconceptions tested at this stage (preselected on the basis of a pilot study with an independent sample of 55 teachers, see Materials below) were to be the subject of the manipulations conducted during Phase 2. Average completion time for Phase 1 was approximately 10 min.

Between Phase 1 and Phase 2 there was a fixed delay of 45 days. During Phase 2, participants were exposed to three types of refutation texts: text along with discrediting information about the origin of the misinformation (text-and-origin; TO), text alone (TA), and no text (NT). Orthogonally, the refutation texts addressed misconceptions of three different levels of endorsement strength, so that the type of text per topic and misconception endorsement level could be counterbalanced. That is, for each level of endorsement (high, medium, and low) one misconception was refuted with a text that contained information about the origin of the misconception (TO), another misconception was refuted with a text without such information (TA), and another misconception was left unchallenged (NT). Therefore, each participant read six refutation texts in total. Texts were presented in a random order for each participant. Immediately after reading the texts, participants completed the same questionnaire used in Phase 1 for a second time. Finally, participants were asked to rate how difficult they found to read the refutation texts on a scale from 1 to 10. Average completion time for this task was approximately 20 min. All the materials were presented on-line.

Materials

Phase 1. The questionnaire consisted of two parts. The first part contained a consent form and requested background information about the participants: Age, sex, professional

qualification (e.g., degree, master degree, PhD), years of teaching, level at which they teach (e.g., kindergarten, elementary school, secondary school), position within the school (e.g., teacher, coordinator, headmaster), type of school they attended (e.g., public school, private school), and the region where their school was located. For exploratory reasons we also collected additional information that we have considered in previous research on misconceptions among teachers (interest in neuroscience applied to education, in-service training in educational neuroscience, regular reading of magazines about general or educational science and/or peer-review journal articles, and regular consultation of books, blogs, or websites about neuroscience).

The second part of the questionnaire contained 36 statements about education and neuroscience applied to education. Half of them are backed by robust evidence and the remaining 18 statements have null or very weak evidence and can be considered misconceptions (see Tables 1 and 2). These items were extracted from previous questionnaires (Dekker et al., 2012), meta-analyses (Ferrero, West, & Vadillo, 2017; Hattie, 2009; Leong, Carter, & Stephenson, 2015), unsystematic reviews (American Academy of Pediatrics, 2011; Bangerter & Heath, 2010; De Bruyckere, Kirschner, & Hulshof, 2015; Geake, 2008; Hyatt, Stephenson, & Carter, 2009; Muijs & Reynolds, 2011; Waterhouse, 2006), experimental studies (Neuman, Kaefer, Pinkham, & Strouse, 2014), international classification and diagnostic manuals (American Psychiatric Association, 2013; World Health Organization, 2016), and national reports (National Reading Panel, 2000). For each question, a 5-point Likert scale was employed, labelled as (1) Definitely false, (2) False, (3) Don't know, (4) True, and (5) Definitely true, indicating the degree of endorsement.

Phase 2. For the purpose of the experiment, we addressed nine of the 18 misconceptions included in the 36-item questionnaire. As explained above, these nine misconceptions differed in terms of their strength ratings derived from the pilot study: Three

misconceptions were strongly endorsed by participants in the pilot study, three had an intermediate level of endorsement, and three were weakly endorsed. The mean endorsement strength and standard deviation of these items in the pilot study are shown in Table 2 (see Appendix 1).

For each misconception, we created one refutation text with three different versions: (a) refutation text which discredited the origin of the misinformation (text-and-origin); (b) refutation text alone (TA); (c) no text (NT) (see the SM). All the texts were structured as follows: At the beginning, the target misconception was introduced and, immediately afterwards, it was refuted. Next, the origin of misinformation was discredited (only in the text-and-origin condition). Then, the alternative (and correct) information was presented. Finally, a rhetorical question was formulated. The mean length of texts which contained the origin of information was 180.66 words (SD = 6.11) and the mean length of the remaining texts was 149.55 (SD = 7.87). After the refutation texts, we included a question aimed at measuring the level of difficulty of the refutation texts according to the perception of the participants. The Likert scale employed for this question ranged from 1 (Extremely easy) to 10 (Extremely difficult). The materials used in both phases are publicly available at https://osf.io/5d6nz/.

Results and Discussion

Figure 1 shows participants' endorsement ratings separated by condition and time. As can be seen, overall participants' endorsements for each misconception declined from Phase 1 to Phase 2. However, this change was steeper for misconceptions that had been addressed with a refutation text. The figure suggests that this decline was roughly similar for the text-and-origin and text-alone conditions.

We analysed the data using a linear mixed effects model[1] with condition (3 levels: TO, TA, NT) and time (2 levels: Phase 1 and 2) as fixed effects, and participant-specific random intercepts. The analysis revealed that both main effects were statistically significant (condition: $\chi^2(2) = 54.86$, $p < .001$; time: $\chi^2(1) = 233.70$, $p < .001$). The effectiveness of refutation texts was assessed through the interaction between time and condition, which was statistically significant, $\chi^2(2) = 42.56$, $p < .001$. Following up the interaction with post-hoc pairwise comparisons (Tukey method), we found that while in Phase 1 there were no differences between conditions (all ps > .05), in Phase 2 both intervention conditions (TO and TA) were significantly different from the NT condition. These results suggest that, relative to the NT condition, the presentation of a simple refutation text (TA) or a refutation text addressing the origin (TO) gave rise to a stronger change in participants' endorsement for misconceptions. Including discrediting information about the origin of the misinformation did not make a significant different in the effect of refutation texts, as the contrast between TO and TA was not significant, $t(760) = 1.04$, $p = .55$.[2] Adding years of experience in the model did not provide a better fit.

Another interesting aspect of the data that warrants further exploration is whether the degree of endorsement in Phase 1 (i.e., the pre-existing strength of each misconception) had an effect on the change in endorsement between Phases 1 and 2. To test this hypothesis, we used a linear mixed effects model predicting change (difference in ratings between Phases 2 and 1, with negative scores indicating lower degree of misconception endorsement in Phase 2) by Phase 1 rating and condition (fixed effects), and random intercepts for each participant.[3] The analysis showed that the higher ratings of endorsement in Phase 1 led to larger changes in Phase 2, $b = -0.63$, $t(376.10) = -13.57$, $p < .001$. This result only holds for the refutation text conditions and not the NT condition, as indicated by a significant interaction, $\chi^2(2) = 35.01$, $p < .001$. Specifically, the interaction contrast that includes a comparison between TA

and TO was not significant, b = -0.08, t(365.10) = -1.39, p = .17, while the interaction contrast that includes NT was significant, b = -0.19, t(366.50) = -5.72, p < .001. In other words, after accounting for the strength of each misconception in Phase 1, we find that refutation texts have an effect on behaviour compared to NT, but the effect of different types (TO vs TA) of intervention is indistinguishable.

The results showed that refutation texts were effective at mitigating endorsement of misconceptions. Specifically, when teachers were presented with misconceptions within refutation texts, they significantly reduced their belief in these ideas in comparison with the misconceptions that were not refuted. In contrast, the data showed that providing discrediting information about misconception origin did not result in a significant additional reduction in the endorsement of misconceptions: Refutation texts including this information (text-and-origin) were not significantly more effective than refutation texts that did not include it (text-alone). Contrary to our initial hypothesis, the most strongly endorsed misconceptions among educators were no more resistant to change than the weaker ones. If anything, we observed a trend in the opposite direction: Participants' change of attitudes in response to the refutation text was largest for the strongest misconceptions. Note, however, that these analyses necessarily conflate strength of belief and item: The items that were most strongly endorsed by our participants at Phase 1 might have some particular feature that makes them more susceptible to correction attempts. Similarly, this pattern might be entirely due to a floor effect among the least popular items, for which the rating scales provide little space for additional correction.

Experiment 2

The results of Experiment 1 suggest that refutation texts might be an effective means to correct erroneous beliefs about education among in-service teachers. However, based just on

this limited evidence, we cannot know whether the brief exposure to a refutation text had long-term effects on participants' beliefs. Furthermore, the fact that participants were less willing to endorse educational misconceptions after reading the refutation texts does not necessarily mean that they will be less likely to rely on instructional methods that are based on those misconceptions. In fact, previous research in the domain of attitudes towards vaccination shows that attempts to correct misinformation might actually make parents less willing to vaccinate their children, even if they also show less trust in the accuracy of the misinformation (Nyhan & Reifler, 2015; Nyhan et al., 2014). To address these concerns, Experiment 2 included new measures to assess whether the manipulation had any effect on teachers' intent to apply specific educational practices the classroom. In addition, to test the persistence of these changes over time, we included an additional test phase 30 days after completion of Phase 2. Finally, it is possible that the impact of discrediting misconception origin in Experiment 1 was reduced due to participants ignoring this information. In an effort to address this in Experiment 2, we highlighted this information with bold text.

Method

Participants

As in Experiment 1, our intention was to test the maximum number of participants that we could reach using the same recruitment strategy as in Experiment 1. Before testing any participant, we conducted a power analysis to decide what would be the minimal sample size that would provide reasonable power to reproduce the main findings of Experiment 1. In the previous experiment, the slope coefficient of the crucial target condition × time interaction was $b = 0.189$. Using the 'simr' package, we were able to estimate that we would need to test at least 23 participants to replicate an interaction with a somewhat smaller slope coefficient, b

= 0.15, with 85% power. Therefore, we decided a priori that we would test at least 23 participants, although we would try to include more participants in our sample if possible. In the end, we were able to test a total of 35 teachers (26 female) from three different schools in Spain. The mean age of participants was 40.91 (*SD* = 9.88). Participants were kindergarten teachers (22.85%), primary school teachers (37.14%), and secondary school teachers (40%). The sample was recruited from public (8.57%) and state schools (91.42%). The average teaching experience of participants was 16.25 years (*SD* = 8.89).

Design and Procedure

The design and procedure were identical to Experiment 1 with three exceptions. Firstly, the piece of text discrediting the original claim behind the misconception was presented in bold characters. Secondly, after reading the six refutation texts and completing the 36-items on-line questionnaire in Phase 2, participants responded to a second on-line questionnaire of 18 items assessing whether they were planning to implement a number of practices in the classroom. Nine of these practices were actually based on the nine crucial misconceptions preselected for the experimental manipulation (see Materials). Thirdly, we included a new test at Phase 3, 30 days after Phase 2. During Phase 3, participants completed the same questionnaires used in Phase 2, but they did not read the six refutation texts nor respond the question about their level of difficulty. For exploratory purposes, Phase 3 included two extra questions per item to determine whether participants had searched any additional information about the beliefs addressed in the refutation texts since Phase 2 and whether they had received any formal training regarding about them.

Materials

Phase 1. The materials were identical to Experiment 1.

Phase 2. The refutation texts were identical to those of Experiment 1, except that the sentences explaining the origin of misconceptions in the text-and-origin condition were now presented in bold characters. In addition to the materials employed in Experiment 1, we created a new survey with 18 educational practices based on each of the 18 statements extracted from the 36-items questionnaire. Half of them were elaborated from the items addressed in the refutation texts, and hence had very weak evidence, and the other half were based on robust empirical evidence. For each question, participants were asked to report how likely they were to transfer each practice into the classroom, using a 6-point Likert scale, labelled as (1) Definitely not, (2) Very unlikely, (3) Unlikely, (4) Likely, (5) Very likely, and (6) Definitely yes.

Phase 3. Participants were asked to fill again the 36-item questionnaire of Phases 1 and 2 and the 18-educational practices survey of Phase 2. Additionally, we added two questions for each of these 18 educational practices. One was aimed at determining if participants had searched for any additional information about each of the beliefs targeted in the refutation texts during the 30 days between Phase 2 and Phase 3. For each belief, there were four response options: (1) I have not searched for information; (2) I do not remember having searched for information; (3) I have searched for information and it runs in the same direction of the refutation text; (4) I have searched for information and it runs in the opposite direction of the refutation text. The second question was aimed at assessing whether participants had received any information regarding each of these beliefs during their academic training. The response options were identical, except that "search for" was replaced by "received".

Results and Discussion

Analysis of the data in Experiment 2 was very similar to that of Experiment 1. Figure 2 (left panel) shows the endorsement rates across the three different experimental phases:

While there is a clear decrease in the mean endorsement in Phase 2 for the refutation text conditions (TO, TA), replicating the effect observed in Experiment 1, there seems to be an increase in Phase 3, almost reaching the mean endorsement before refutation texts were presented (i.e., Phase 1).

For this analysis, we used a linear mixed effects model with condition (3 levels: TO, TA, NT) and time (3 levels: Phase 1, 2, and 3) as fixed effects, and participant-specific random intercepts. All main effects were significant (condition: $\chi^2(2) = 32.74$, $p < .001$; time: $\chi^2(2) = 155.41$, $p < .001$). As in Experiment 1, the significant interaction ($\chi^2(4) = 28.86$, $p < .001$) showed that there was no evidence of a difference between conditions in Phase 1, however, in Phase 2 both experimental conditions were significantly different from the NT condition (all *ps* < .01), but not between each other, $t(926) = 1.22$, $p = .44$. Interestingly, the mean endorsement for misconceptions increased in Phase 3 where all pairwise differences between the 3 conditions were non-significant (all *ps* > .13).

In this experiment, we also asked participants to rate the difficulty of the refutation texts. The new model showed that perceived difficulty had a significant effect on the rate of endorsement, $b = 0.14$, $t(294.89) = 2.30$, $p = .02$, indicating that as difficulty increased, so did the mean rate of endorsement, regardless of condition and Phase (i.e., all 2-way and 3-way interactions including difficulty were not reliable, all *ps* > .15). In contrast, adding years of experience in the model did not provide a better fit (likelihood-ratio test, $\chi^2(18) = 14.84$, $p = .67$).

We conducted the same type of analysis for the intention-to-use statements. Figure 2B shows a clear pattern that across Phases 2 and 3 the refutation text conditions received higher ratings of intention-to-use. This is a rather surprising result considering the mean endorsement rate from the previous analysis. The main effect of condition was significant ($\chi^2(2) = 98.87$, $p < .001$), with both intervention conditions attaining higher intention to use

rates than the NT condition (both *ps* < .001). The effect of time was significant with participants giving higher ratings in Phase 3 than Phase 2, $\chi^2(1) = 3.96$, $p = .046$), but the interaction between time and condition was not significant, $\chi^2(4) = 5.93$, $p = .051$. Introducing difficulty and years of experience in the main model did not result in better predictive power (likelihood-ratio tests, both *ps* > .05.

Finally, we also conducted two extra analyses addressing the change in endorsements between phases. Both linear mixed-models modelled the change in endorsements between Phases (model 1: difference in ratings between Phase 2 and Phase 1; model 2: difference in ratings between Phase 3 and Phase 2) as predicted by the endorsement rating in the preceding Phase (i.e., rating in Phase 1 for model 1 and rating in Phase 2 for model 2). In both models, we found that higher ratings in the previous Phase led to larger differences in the following Phase (model 1: $b = -0.28$, $t(303.82) = -3.44$, $p < .001$; model 2: $b = -0.32$, $t(308.54) = -4.41$, $p < .001$). This result is moderated by condition between Phases 2 and 1 (model 1: $\chi^2(2) = 16.86$, $p < .001$) for the refutation text conditions, but not between Phases 3 and 2 (model 2: $\chi^2(2) = 2.64$, $p = .27$).

The present results largely replicate the main findings of Experiment 1, but with important caveats. As in Experiment 1, participants were more reluctant to endorse misconceptions after reading the refutation texts and this effect was, if anything, larger for strongly held beliefs. Again, addressing the origins of misconceptions, even using bold characters for this piece of information, made no significant difference and the effects were not moderated by years of experience, either. However, the additional tests included in Experiment 2 reveal that the effects of refutation texts were short-lived and vanished after a 30-day period. Furthermore, the manipulation did not have a beneficial effect on intention-to-use ratings, where, in fact, a small but significant backfire effect was found.

General Discussion

Teachers all over the world hold a significant number of misconceptions related to education (i.e., Dekker et al., 2012; Ferrero et al., 2016). This fact might be damaging for the school population because it facilitates the adoption of ill-funded interventions and, consequently, jeopardises the quality and rigor of educational practice. In spite of this, no systematic study to date has evaluated different strategies to correct these false beliefs among in-service teachers. In the present study, we exposed a group of teachers to different types of refutation texts in order to ameliorate the acceptance of misconceptions about education and neuroscience applied to education. The key elements of refutation texts are the introduction of a popular belief or idea, its immediate refutation, and the presentation of a more satisfactory alternative to replace the previous one (Guzzetti, 2000).

In the present experiments, when teachers were presented with misconceptions corrected within refutation texts, they significantly reduced their belief in these ideas in comparison with the misconceptions that were not refuted. These results are consistent with previous studies aimed at testing the success of refutation texts to correct misconceptions in different fields, such as health or specific academic subjects (i.e., Guzzetti et al., 1997; Hynd et al., 1994; Nyhan et al., 2014), although some studies also detected a continued influence effect which impeded the full effectiveness of these texts (i.e., Nyhan & Reifler, 2015). Interestingly, providing additional information that discredited the origin of misconceptions did not enhance the effects of refutation texts significantly in either experiment, even when this additional information was highlighted using a bold font. This result stands in contrast with previous research showing that suspicion about the reliability of the original misinformation or its source might foster belief change (Guillory & Geraci, 2013; Lewandowsky et al., 2005).

Perhaps the most striking feature of our results is that the effects of refutation texts were short-lived. We failed to detect a significant effect of the manipulation after a relatively short period of 30 days. These results are consistent with previous studies that find no permanent effect of refutation texts (Gregg, Winer, Cottrell, Hedman, & Fournier, 2001). Similarly, we found no significant evidence of changes in participants' intention to implement educational practices that are based on the misconceptions addressed in the refutation. If anything, attempts to correct misconceptions had a paradoxical effect on intention-to-use ratings: Participants were actually more willing to implement educational practices based on the misinformation after reading the refutation texts. Although unexpected, these results converge with previous research on the 'backfire' effect showing that correcting misconceptions about a specific topic does not necessarily result in an increase in the intention to implement correct practices. For instance, among parents hesitant about vaccination, attempts to correct misinformation might work at the knowledge level, but be counterproductive when it comes to changing intention to vaccinate (i.e., Nyhan & Reifler, 2015; Nyhan et al., 2014).

Contrary to our initial hypothesis, the most strongly endorsed misconceptions among educators were no more resistant to change than the weaker ones. If anything, we observed a trend in the opposite direction: Participants' change of attitudes in response to the refutation text was largest for the strongest misconceptions. As mentioned in the introduction, previous studies have generally found that firm beliefs are more difficult to change (Lewandowsky et al., 2012; Nyhan & Reifler, 2015), but the opposite result has been observed as well (Butterfield & Metcalfe, 2001; see also Ecker, Lewandowsky et al., 2014; van Loon et al., 2015). It is possible that, at least in some domains, attempts to correct firmly held beliefs are more effective at drawing participants' attention to the error or simply enhance error-correction processes in the manner predicted by classic associative learning models (Rescorla

& Wagner, 1972). In the same vein, the number of years as in-practice teacher did not moderate the effectiveness of refutation texts.

Of course, the present study is not without limitations. As in any experiment, our results might not generalize beyond these specific materials and samples. In the same vein, although we tried to test as many participants as possible, this is a relatively difficult to reach population, which limited our sample sizes to just 45 and 35 participants in Experiments 1 and 2, respectively. We cannot discard the possibility that our failure to find significant effects for some manipulations was simply due to the lack of statistical power. As a first attempt to evaluate the effectiveness of this intervention with in-service teachers, we wanted to use experimental conditions that would maximize the chance of observing a change in endorsement. Consequently, we decided to compare both experimental conditions (text-and-origin and text-alone) with a no-text condition. However, once the effectiveness of refutation texts is firmly established, future research should consider the introduction of a non-refutation text control condition, addressing only the correct information without explicit mention of the misconception. It would also be informative to directly measure behaviour, instead of measuring intention to implement misguided practice. Finally, none of our experiments were formally preregistered. We acknowledge that this might introduce biases in our analytic pipeline that, ideally, should be controlled for in future studies.

As mentioned before, the high prevalence of misconceptions among teachers all over the world is well documented. These false ideas are often introduced in schools through workshops, conferences, and educational materials and pave the way for the use of pseudoscientific practices in classroom (Busso & Pollack, 2014; Goswami, 2006). In the face of this worrying reality, some experts have recommended efforts to improve the knowledge of educators about specific topics and research methodologies (Ansari, Coch, & De Smedt, 2011; Goswami, 2004; Lilienfeld, Admmirati, & David, 2012), to strengthen the

collaboration between researchers and teachers (Ansari et al., 2011; Howard-Jones, 2014), or to explicitly address the most popular misconceptions among educators (Ferrero et al., 2016). In the present study we found that refutation texts only had a short-lived mitigating effect on common misconceptions about education, and had the unintended side-effect of increasing intention to implement instructional methods based on misconceptions. More research is needed to find alternative methods that ameliorate misconceptions in the long-term and facilitate transfer to teachers' intention not to use questionable educational methods.

Open practices statement

None of the present studies were preregistered. In addition to the analyses presented here, we also analyzed the data from Experiment 1 with ANOVA models, which yielded virtually identical results and are omitted for the sake of simplicity. All data, materials, and analysis scripts related to this study will be publicly available at https://osf.io/73a9y/. All data exclusions, manipulations, and measurements conducted during this study are reported in the Methods sections of Experiments 1 and 2.

References

Ansari, D., Coch, D., & De Smedt, B. (2011). Connecting education and cognitive neuroscience: Where will the journey take us? *Educational Philosophy and Theory, 43*, 37-42.

American Academy of Pediatrics (1998). Auditory integration training and facilitated communication for autism. *Pediatrics, 102*, 431-433.

American Academy of Pediatrics (1999). The treatment of neurologically impaired children using Patterning. *Pediatrics, 104*, 1149-1151.

American Academy of Pediatrics (2011). Learning disabilities, dyslexia, and vision. *Pediatrics, 124*, 837-844.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

American Speech-Language-Hearing Association (n.d.). [Auditory integration training]. Retrieved from https://www.asha.org/policy/tr2004-00260.htm

Bangerter, A., & Heath, C. (2010). The Mozart effect: Tracking the evolution of a scientific legend. *British Journal of Social Psychology, 43*, 605-623.

Busso, D. S., & Pollack, C. (2014). No brain left behind: Consequences of neuroscience discourse for education. *Learning, Media and Technology*, DOI: 10.1080/17439884.2014.908908

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491-1494.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences, 14*, 161-199.

Coffield, F., Moseley, D., Hall, E., & Eclestone, K. (2004). *Learning styles and pedagogy in post-16 learning. A systematic and critical review.* London: LSRC & Dept. for Education and Skills.

Corballis, M. C. (1999). Are we in our right minds? In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain.* Chichester, UK: Wiley.

De Bruyckere, P., Kirschner, P. A., & Hulshof, C. D. (2015). *Urban myths about learning and education*. New York: Academic Press.

Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology, 3*, 429.

Deligiannidi, K., & Howard-Jones, P. A. (2015). The neuroscience literacy of teachers in Greece. *Procedia - Social and Behavioral Sciences, 174*, 3909-3915.

Ecker, U.K.H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition, 42*, 292-304.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition, 38*, 1087-1100.

Ecker, U. K. H., Swire, B., & Lewandowsky, S. (2014). Correcting misinformation—A challenge for education and cognitive science. In D. N. Rapp & J. L. G. Braasch (Eds.). *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 13-38). Cambridge, MA: MIT Press.

Ferrero, M., Garaizar, P., & Vadillo, M. A. (2016). Neuromyths in education: Prevalence among Spanish teachers and an exploration of cross-cultural variation. *Frontiers in Human Neuroscience, 10*, 496.

Ferrero, M., West, W., & Vadillo, M. A. (2017). Is crossed laterality associated to academic achievement and intelligence? A systematic review and meta-analysis. *PLoS ONE, 12*:e0183618.

Fuentes, A., & Riso, A. (2015). Evaluación de conocimientos y actitudes sobre neuromitos en futuros/as maestros/as [Evaluation of neuromyth's knowledge and attitudes in future research]. *Revista de Estudios de Investigación en Psicología y Educación, 6*, 193-198.

Freed, G. L., Clark, S. J., Butchart, M. T., Singer, D. C., & Davis, M. M. (2009). Parental vaccine safety concerns in 2009. *Pediatrics, 125*, 654-659.

Geake, J. (2008). Neuromythologies in education. *Educational Research, 50*, 123-133.

Gill, M. G., Ashton, P. T., & Algina, J. (2004). Changing pre-service teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology, 29*, 164–185.

Gleichgerrcht, E., Luttges, B. L., Salvarezza, F., & Campos, A. L. (2015). Educational neuromyths among teachers in Latin America. *Mind, Brain, and Education, 9*, 170-178.

Gregg, V. R., Winer, G. A., Cottrell, J. E., Hedman, K. E., & Fournier, J. S. (2001). The persistence of a misconception about vision after educational interventions. *Psychonomic Bulletin & Review, 8,* 622-626.

Goldacre, B. (2006). Brain Gym - Name & Shame [Blog post]. Retrieved from http://www.badscience.net/2006/03/the-brain-drain/

Goswami, U. (2004). Neuroscience and education. *British Journal of Educational Psychology, 74*, 1-14.

Goswami, U. (2006). Neuroscience and education: From research to practice? *Nature Reviews Neuroscience*, 7, 406-413.

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition, 2,* 201-209.

Guzzetti. B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly, 16*, 89-98.

Guzzetti, B. J., Williams, W. O., Skeels, S. A., & Wu, S. M. (1997). Influence of text structure on learning counterintuitive physics concepts. *Journal of Research in Science Teaching, 34*, 701-719.

Haglin, K. (2017). The limitations of the backfire effect. *Research and Politics, 4*, 1-5.

Handler, S. M., Fierson, W. M., Section on Ophthalmology, Council on Children with Disabilities, American Academy of Ophthalmology, American Association for Pediatric Ophthalmology and Strabismus, American Association of Certified Orthoptists (2011). Joint technical report—Learning disabilities, dyslexia, and vision. *Pediatrics, 127*, 818-856.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London & New York: Routledge.

Hines, T. (1987). Left brain/right brain mythology and implications for management and training. *Academy of Management Review, 12*, 600-606.

Howard-Jones, P. A. (2014). Neuroscience and education: myths and messages. *Nature Reviews Neuroscience, 15*, 817-824.

Hunter, A. S., & Lloyd, M. E. (2018). Faculty discuss strategies, but not the best ones: A survey of suggested exam preparation techniques for difficult courses across disciplines. *Scholarship of Teaching and Learning in Psychology, 4*, 105-114.

Hyatt, K. J. (2007). Brain Gym®: Building stronger brains or wishful thinking? *Remedial and Special Education, 28*, 117-124.

Hyatt, K. J., Stephenson, J., & Carter, M. (2009). A review of three controversial educational practices: Perceptual motor programs, sensory integration, and tinted lenses. *Education and Treatment of Children, 32*, 313-342.

Hynd, C. R. (2001). Refutational texts and the change process. *International Journal of Educational Research, 35*, 699-714.

Hynd, C. R., McWhorter, J. Y., Phares, V. L., & Suttles, C. W. (1994). The role of instructional variables in conceptual change in high school physics topics. *Journal of Research in Science Teaching, 31*, 933-946.

Johnson, J. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1420-1436.

Karakus, O., Howard-Jones, P. A., & Jay, T. (2015). Primary and secondary school teachers´ knowledge and misconceptions about the brain in Turkey. *Procedia- Social and Behavioral Science, 174*, 1933-1940.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 35*, 1567-1577.

Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology, 36*, 153-159.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480-498.

Leong, H. M., Carter, M., & Stephenson, J. R. (2015). Meta-analysis of research on sensory integration therapy for individuals with developmental and learning disabilities. *Journal of Developmental and Physical Disabilities, 27*, 183–206.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106-131.

Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., & Smithson, M. (2015). Climate change denial and its effect on the scientific community. *Global Environmental Change, 33*, 1-13.

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science, 16*, 190-195.

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology, 50*, 7-36.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*, 232-257.

Maria, K., & MacGinitie, W. (1987). Learning from texts that refute the reader's prior knowledge. *Reading, Research and Instruction, 26*, 222-238.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2015). Instructor and student knowledge of study strategies. *Memory, 24,* 257-271.

Muijs, D., & Reynolds, D. (2011). *Effective teaching. Evidence and practice*. London: Paul Chapman.

National Reading Panel (U.S.). (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, D.C.: National Institute of Child Health and Human Development, National Institutes of Health.

Neuman, S. B., Kaefer, T., Pinkham, A., & Strouse, G. (2014). Can babies learn to read? A randomized trial of baby media. *Journal of Educational Psychology, 106*, 815–830.

New York State Department of Health (1999). *New York State Department of Health early intervention program. Clinical practice guideline. Report of the recommendations.* Retrieved from https://www.health.ny.gov/community/infants_children/early_intervention/disorders/autism/app_c.htm

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behaviour, 32*, 303-330.

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine, 33*, 459-464.

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics, 133*, e835-e842.

Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The hazards of correcting myths about health care reform. *Medical Care, 51*, 127-132.

OECD (2002). *Understanding the brain. Towards a new learning science*, OECD Publishing, Paris. Retrieved from: http://www.oecd.org/edu/ceri/31706603.pdf

Pashler, H., McDanield, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Eye, 9*, 105-119.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 2*, 211-227.

Rao, T. S. S., & Andrade, C. (2011). The MMR vaccine and autism: Sensation, refutation, retraction, and fraud. *Indian Journal of Psychiatry*, 53, 95-96.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F.

Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century Crofts.

Salisbury-Glennon, J. D., & Stevens, R. J. (1999). Addressing pre-service teachers' conceptions of motivation. *Teaching and Teacher Education, 15*, 741-752.

Scahill, L., & Bears, K. (2009). The rise in autism and the mercury myth. *Journal of Child and Adolescent Psychiatric Nursing, 22*, 51-53.

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1948-1961.

Tippet, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education, 8*, 951-970.

Van Loon, M., Dunlosky, J., van Gog, T., van Merriënboer, J. J.G., & de Bruin, A. B. H. (2015). Refutation in science texts lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology, 42*, 39-45.

Vergara, G., Martínez, M., Martínez-Sahuquillo, M. E., & Echevarría, C. (2011). Eficacia del método de los Institutos para el Logro del Potencial Humano (Doman-Delacato) en pacientes con parálisis cerebral infantil [Eficcacy of the method of the Institutes for the Achievement of Human Potential (Doman-Delacato) in patients with infant cerebral palsy]. *Rehabilitación, 45*, 256-260.

Waterhouse, L. (2006). Multiple Intelligences, the Mozart effect, and emotional intelligence: A critical review. *Educational Psychologist, 41*, 207-225.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology, 40*, 361-387.

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behaviour, 41*, 135-163.

World Health Organization (2016). *International statistical classification of diseases and related health problems (10th rev.).* Geneva: World Health Organization.

Author Note

MF was supported by grant AYD-000-235 from bizkaia:talent, Diputación Foral de Bizkaia, and by a postdoctoral grant from Programa Posdoctoral de Perfeccionamiento de Personal Investigador Doctor, Gobierno Vasco. The Meta-Research Innovation Center at Stanford (METRICS) is supported by a grant from the Laura and John Arnold Foundation. The Meta-Research Innovation Center Berlin (METRIC-B) is supported by a grant from the Einstein Foundation and Stiftung Charitè. MAV was supported by grants 2016-T1/SOC-1395 (Comunidad de Madrid) and PSI2017-85159-P (AEI/FEDER UE). TH was based at METRICS when the study began and is now based at METRIC-B**.** The authors are grateful to all the teachers who contributed to the completion of this research. Correspondence concerning this article should be addressed to Marta Ferrero, Facultad de Educación, Universidad Complutense de Madrid, Madrid, Spain, or to Miguel A. Vadillo, Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain E-mail: maferr16@ucm.es or miguel.vadillo@uam.es

Footnotes

[1] All analyses reported in the text were run in R using the packages lme4 for mixed effects models and emmeans.

[2] Due to the fact that the dependent variable is ordinal, we tried a cumulative link mixed effect models (CLMM) analysis, which provided the same pattern of results as the linear mixed effects model presented above.

[3] This analysis is structurally equivalent to an ANCOVA, where the effect of condition on the change between Phases 1 and 2 is tested after accounting for the degree of endorsement in Phase 1.

Table 1. *Items with robust evidence.*

1. Direct instruction leads to better outcomes than discovery learning.
2. Reversing letters is NOT a symptom of dyslexia.
3. The impact of new technologies on learning is questionable.
4. Boys have bigger brains than girls.
5. It is more effective to give feedback that includes information about the correct answer than information about the incorrect answer.
6. Accelerated learning is an appropriate approach for gifted students.
7. The proportion of boys diagnosed with autism is greater than the proportion of girls.
8. Normal development of the human brain involves the birth and death of brain cells.
9. Extended rehearsal of some mental processes can change the shape and structure of some parts of the brain.
10. Phonological knowledge, alphabetic principle, fluency, vocabulary, and comprehension are key elements in the teaching of reading.
11. The left and right hemispheres of the brain always work together.
12. Production of new connections in the brain can continue into old age.
13. Homework is more beneficial for secondary students than for elementary students.
14. We use our brains 24 h a day.
15. Spacing practice is more effective than concentrating the same amount of practice in a shorter period of time.
16. Differences between boys and girls are negligible when teaching one another.
17. Information is stored in the brain in a network of cells distributed throughout the brain.
18. To diagnose a child with autism, symptoms should be present in two or more contexts (i.e. at home and at school).

Table 2. *Items with null or very weak evidence and their degree of endorsement on the basis of the pilot study.*

Item	Mean (*SD*)	Reference
1. Environments that are rich in stimulus improve the brains of pre-school children. (*)	4.58 (0.59)	Goswami (2004)
2. Individuals learn better when they receive information in their preferred learning style (e.g., auditory, visual, and kinesthetic). (*)	4.50 (0.80)	Coffield, Moseley, Hall, & Eclestone (2004)
3. Adapting teaching methods to the "multiple intelligences" of students leads to better learning.	4.47 (0.65)	Geake (2008)
4. Exercises that rehearse co-ordination of motor-perception skills can improve literacy skills. (*)	4.43 (0.70)	Hyatt et al. (2009)
5. Short bouts of coordination exercises can improve integration of left and right hemispheric brain function.	4.20 (0.77)	Hyatt (2007)
6. Children with learning difficulties and autism can benefit from controlled sessions of sensorial stimulation (i.e., being swayed in an hammock or be stroked with a brush).	4.14 (0.96)	Leong et al. (2015)
7. Differences in hemispheric dominance (left brain, right brain) can help explain individual differences among learners. (*)	3.76 (0.95)	Geake (2008)
8. Several reading difficulties can be remediated by optometric exercises (i.e., to follow the movement of a ball with the eyes, to number letters line by line using exclusively the eyes).	3.65 (1.04)	Handler et al. (2011)
9. New generations of students have sophisticated technical skills that helps them to learn more efficiently using information from the Internet. (*)	3.63 (0.86)	De Bruyckere et al. (2015)
10. We only use 10% of our brain. (*)	3.60 (1.35)	Geake (2008)
11. Learning and behaviour disorders can be reduced by short bouts of listening to electronically modified music (i.e., Berard method).	3.41 (0.84)	American Academy of Pediatrics (1998)
12. The number of children with autism has increased significantly over the last years.	3.25 (0.81)	Scahill, & Bears, (2009)
13. Listening to the classical music of Mozart enhances childrens' intelligence. (*)	3.25 (0.93)	Waterhouse (2006)
14. Crossed laterality (i.e., right hand dominant and left eye dominant) is associated with learning disabilities. (*)	3.21 (1.30)	Ferrero et al. (2017)
15. The whole-language approach (teaching whole words and not the correspondence letter-sound) is most appropriate in the early teaching of reading. (*)	3.16 (1.24)	National Reading Panel (2000)
16. There are critical periods in childhood after which certain things can no longer be learned.	2.58 (1.07)	OECD (2002)
17. The measles-mumps-rubella vaccine might cause autism.	2.34 (1.04)	Rao & Andrade, (2011)
18. The majority of babies (3-18 months) can learn to read using an appropriate method.	2.34 (1.06)	Neuman et al. (2014)

(*) Misconceptions addressed in Phase 2.

Figure Captions

Figure 1. Participants' endorsement for the misconceptions before (Phase 1) and after reading the refutation texts (Phase 2) in each condition in Experiment 1. TO, TA, and NT refer to Text + Origin, Text Alone, and No Text, respectively. Error bars denote 95% CIs.

Figure 2. Participants' endorsement (left panel) and intention-to-use ratings on Phases 1-3 of Experiment 2. TO, TA, and NT refer to Text + Origin, Text Alone, and No Text, respectively. Error bars denote 95% CIs.

Appendix 1

To determine the popularity of the 18 misconceptions contained in the on-line 36-items survey used in Phase 1 and Phase 2, we conducted a pilot study with 55 (35 female) in-service teachers of 10 Spanish schools, different from the sample who participated in the main study. The items included in this pilot study and the rating scales (1-5) were identical to those described in the main study of the present article (see Table 1 and 2). We recruited educators by a direct invitation published on Twitter containing a link to a questionnaire with the same 36 questions included in the main study. The mean age of participants was 40.10 years (*SD* = 10.70). Participants were special education teachers (12.72%), kindergarten teachers (18.18%), primary school teachers (34.54%), secondary school teachers (24.44%), vocational education teachers (14.54%), and teachers who worked in more than one level of education (3.63%). The sample was recruited from public (45.45%), private (5.45%), and state schools (49.09%). The average teaching experience of the educators was 14.34 years (*SD* = 9.81). Once the task had been completed, we extracted three misconceptions with a high level of endorsement, three with medium endorsement, and three with low endorsement. These nine erroneous ideas were used in the nine refutation texts presented in Phase 2. The results of the pilot study are shown in Table 2.

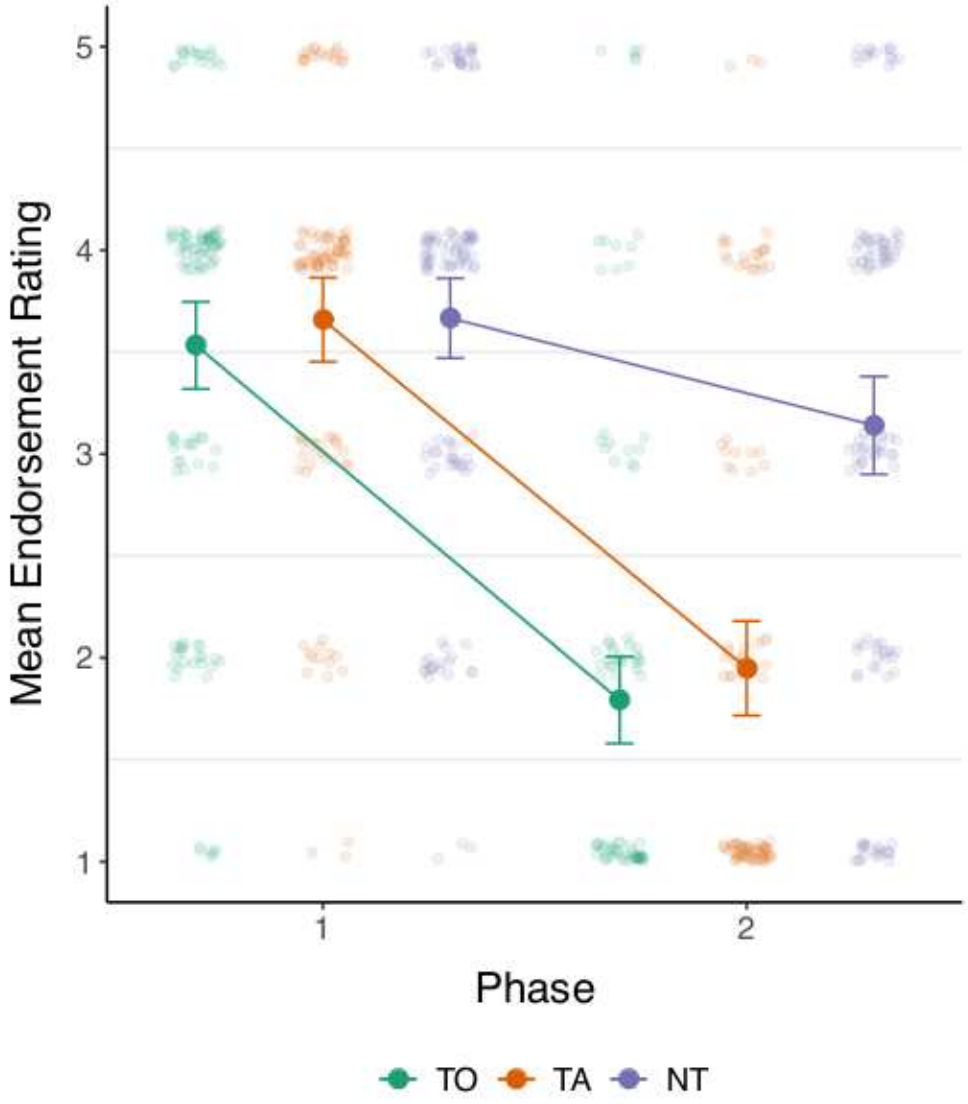

Figure #1

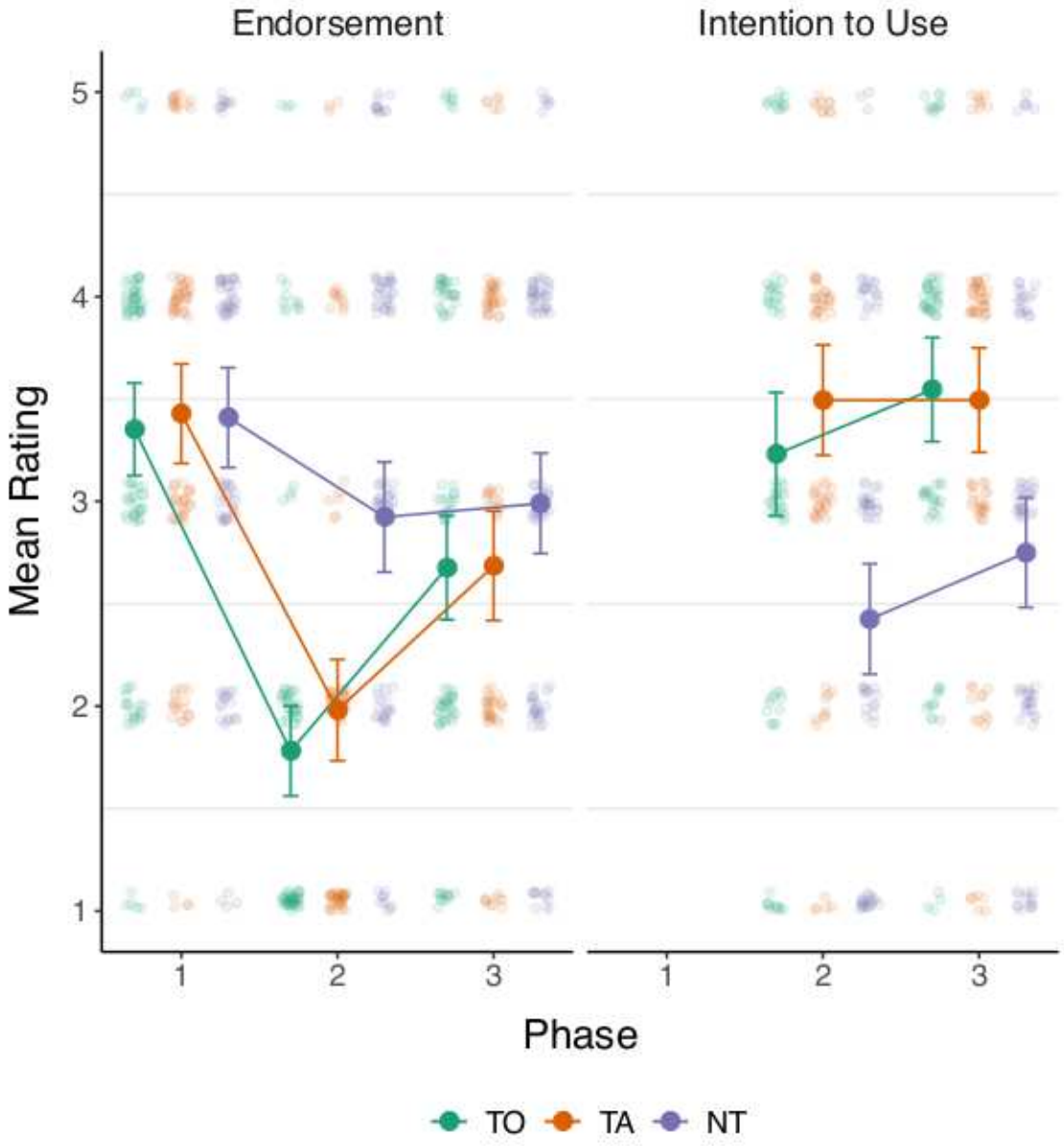
Endorsement
Intention to Use
Mean Rating
5
4
3
2
1
1
2
3
1
2
3
Phase
TO
TA
NT

Figure #2

Supplemental Material

Spanish versions of the nine refutation texts used during Phase 2 of the study and an English approximate translation of them are shown below. Sentences in bold contain the explanation added in the TO condition about the origin of information.

Text 1

There is a common belief that environments that are rich in stimuli improve the development of neuronal connections that enhance intelligence in pre-school children. However, this belief is incorrect. **It is based on the results of rats reared in extreme sensorial deprivation, far removed from the conditions under which a child grows up.** Neurologists show that, except for extreme cases, the formation of connections occurs outside of environmental stimulation. On the one hand, experts show that the formation of neuronal connections begins even before birth, this is, before children receive stimulation from the environment. On the other hand, many studies have found that, in normal conditions, stimuli from the environment do not cause a large formation of neuronal connections even after birth. Moreover, who says that having more neuronal connections (synapses) is better? Several disorders demonstrate that having particularly a high number of synapses is not desirable. People with Fragile-X syndrome is a clear example.

Texto 1

Existe la creencia popular de que los entornos ricos en estímulos incrementan las conexiones cerebrales y por tanto la inteligencia de los niños pre-escolares. Sin embargo, esta creencia es errónea. **Se basa en los resultados con ratas criadas en condiciones de deprivación sensorial extrema, muy alejadas de las condiciones en las que crece un niño humano.** Los neurólogos demuestran que, salvo en casos extremos, la formación de conexiones en el cerebro ocurre al margen de la estimulación ambiental. Por un lado, los expertos muestran que la formación de conexiones neuronales comienza incluso antes del nacimiento, es decir, antes de que los bebés reciban ninguna estimulación del ambiente. Por otro lado, muchos estudios encuentran que, en condiciones normales, después del nacimiento los estímulos ambientales tampoco provocan una mayor formación de conexiones neuronales. Más aún, ¿quién ha dicho que tener más conexiones neuronales (sinapsis) sea mejor? Muchos trastornos evidencian que tener niveles de sinapsis especialmente altos no es deseable. Un claro ejemplo de esto son las personas con el síndrome de X-frágil.

Text 2

The belief that individuals learn better when they receive information in their preferred learning style (e.g., auditory, visual, and kinesthetic) has become very popular. However, it must be noted that this is false and is considered a myth by OECD. **It arose in 1975 when a psychologist popularized a classification of students according to four learning styles, simply on the basis of his personal experience and not on empirical studies.** None of the numerous reviews performed have found that adapting teaching to the preferred

---

learning style benefits students. In addition, classifying a student within a unique style is not always possible and the available tools to do so lack reliability and validity. This does not mean that the same type of instruction is effective in all contexts, subjects and students. But there is a significant gap between this general idea and deciding what type of instruction a student needs based on the commercialised programmes inspired by learning styles. Moreover, on the assumption that the theory is valid, would it be possible to personalize instruction for all students taking into account that there are more than 71 classifications of learning styles on the market?

Texto 2

La creencia de que las personas aprenden mejor cuando reciben la información en su estilo de aprendizaje preferido se ha hecho muy popular. Ahora bien, conviene decir que es falsa. **Se originó en 1975 cuando un psicólogo popularizó una clasificación de los estudiantes en función de cuatro estilos de aprendizaje basándose simplemente en su experiencia personal y no en estudios empíricos.** Ninguna de las numerosas revisiones realizadas encuentra que adaptar la enseñanza al estilo de aprendizaje preferido beneficie a los alumnos. Además no siempre es posible clasificar a un estudiante dentro de un solo estilo y las herramientas disponibles para hacerlo carecen de fiabilidad y validez. Esto no significa que el mismo tipo de instrucción valga en cualquier contexto y materia y para todos los alumnos. Pero de ahí a supeditar la decisión de qué tipo de instrucción necesita cada alumno a los comercializados estilos de aprendizaje hay una brecha importante. Incluso, suponiendo que la teoría fuera válida, ¿sería posible personalizar la enseñanza para todos los alumnos considerando que existen más de 71 clasificaciones de estilos de aprendizaje en el mercado?

Text 3

Programs that promote perceptuo-motor coordination exercises to improve literacy skills have become very popular. However, these programs are ineffective. **They originate from well-known commercial products such as Brain Gym® or Primitive reflexes inhibition therapy whose authors, who are making great fortunes, have been strongly discredited by education and health authorities.** These programs state that doing specific physical exercises can modify the structure of the brain and improve the acquisition of reading, among others. However, cognitive psychology has shown that improving perceptuo-motor skills does not cause the improvement of other higher-order processes such as reading. If this were so, wouldn't it be difficult to find children with dyslexia who are very good at sports? Undoubtedly, this is not the case. Psycholinguistics have shown that complex skills such as reading only improve through direct and specific training. Thus, reading programs that focus on phonological knowledge, grapheme-phoneme correspondence, fluency, vocabulary and comprehension have been shown to be effective at improving reading in several empirical studies.

Texto 3

Los programas que promueven hacer ejercicios de coordinación perceptivo-motora para mejorar la lecto-escritura se han hecho muy populares. Sin embargo, estos programas son ineficaces. **Su origen está en conocidos productos comerciales como el Brain Gym® o la terapia de inhibición de reflejos primitivos cuyos autores, que están amasando grandes fortunas, han sido desacreditados duramente por las autoridades educativas y sanitarias.** Estos programas defienden que realizar determinados ejercicios físicos puede modificar la estructura del cerebro y mejorar así el aprendizaje lector, entre otros. Sin embargo, la psicología cognitiva muestra que mejorar las habilidades perceptivo-motoras no causa la mejora de otros procesos superiores como la lectura. Si así fuera, ¿no debería ser

muy difícil encontrar niños disléxicos a los que se les dé muy bien el deporte? Sin duda, no es el caso. Los psicolingüistas evidencian que las habilidades complejas como la lectura sólo mejoran si se entrenan de forma directa y específica. Así, los programas que trabajan el conocimiento fonológico, la unión letra-sonido, la fluidez lectora, el vocabulario y la comprensión sí han mostrado una gran eficacia en la mejora de la lectura en numerosos estudios científicos.

Text 4

There is a common belief that we only use 10% of our brain. However, it should be noted that this is false. **This belief originates from a misinterpretation by the first researchers to study the brain. At that time and with limited resources, they overlooked the function of 90% of this organ.** Nowadays, clinic neurology has shown that if we only used 10% of our brain, we would be in a vegetative state. None of the brain areas can be destroyed without provoking some kind of functional impairment. In fact, losing much less than 90% of our brain as a result of an accident or an illness has already catastrophic consequences for people, such as a loss of speech or memory. In addition, neurophysiological techniques have shown that there is no area in the brain which is never used. Finally, in light of evolutionary theory, would it make sense that our metabolism spent so many resources in a greatly underutilized organ?

Texto 4

La creencia de que sólo usamos el 10% de nuestro cerebro es muy popular en los medios. Sin embargo, conviene subrayar que es falsa. **Todo comenzó tras una interpretación equivocada de las declaraciones de los primeros investigadores en el cerebro. Éstos, por aquel entonces y con escasos medios, afirmaron que aún desconocían la función del 90% de este órgano.** Hoy en día, la neurología clínica demuestra que si sólo usáramos el 10% de nuestro cerebro estaríamos en estado vegetativo. Ninguna área del cerebro puede ser destruida sin dejar en la persona algún tipo de daño funcional. De hecho, la pérdida de áreas muy pequeñas del cerebro a causa de un accidente o de una enfermedad ya tiene consecuencias catastróficas en las personas, tales como la pérdida del habla o la memoria. Las técnicas de exploración neurofisiológica, a su vez, demuestran que no hay ninguna parte del cerebro que no se use nunca. Por último, bajo la lógica de la evolución, ¿tendría algún sentido que nuestro metabolismo hiciera un gasto tan elevado de recursos en un órgano tan infrautilizado?

Text 5

Over the past few years, the belief that differences in hemispheric dominance (left brain, right brain) can help explain individual differences among learners has become popular. However, this idea is incorrect. **It arose from overgeneralization to healthy people of results obtained with epileptic patients who underwent removal of their corpus callosum, a brain structure that unites the two cerebral hemispheres.** At present, brain scans made with neuroimaging techniques show that both hemispheres work together during all cognitive activities. It is true that there is some hemispheric specialisation for specific skills but, at the same time, in a normal brain there are a large number of inter-hemispheric connections. Moreover, there is no evidence that differences among people in terms of creativity, logic or capacity to get excited are reliably linked to processing differences in either hemisphere. In other words, did you know that all people, from the most creative ones to the most analytic ones, employed the two hemispheres continuously?

Texto 5

Desde hace unos años, se ha extendido la idea de que los estudiantes, y las personas en general, pueden ser clasificados como "de cerebro derecho" (persona artística, creativa, emocional) o "de cerebro izquierdo" (persona racional, analítica, lógica). No obstante, esta idea es errónea. **Nació a partir de la sobregeneralización a población sana de los resultados obtenidos con pacientes epilépticos a los que se extirpaba el cuerpo calloso que une los dos hemisferios cerebrales para estudiar sus reacciones.** Actualmente, las exploraciones del cerebro hechas mediante técnicas de neuroimagen muestran que ambos hemisferios trabajan juntos en todas las tareas cognitivas. Sí es cierto que existe cierta especialización hemisférica para determinadas habilidades pero, al mismo tiempo, en un cerebro normal hay una grandísima cantidad de conexiones inter-hemisféricas. Es más, no hay evidencia que demuestre que las diferencias entre personas en términos de creatividad, lógica o capacidad de emocionarse estén relacionadas con diferencias de procesamiento de uno u otro hemisferio. Dicho de otra forma, ¿sabía que todas las personas, desde las más creativas hasta las más analíticas, utilizan ambos hemisferios continuamente?

Text 6

There is a common belief that new generations of students have sophisticated technological skills that enable them to learn more efficiently using information on the web. Nevertheless, this belief is incorrect. **It emerged when in 2011 an advisor and designer of videogames coined the term “digital native” to describe the youth of today on the basis of anecdotes and informal observations.** Both objective indicators and information obtained though structured interviews with young people of the “new digital age” show that they do not have a better knowledge about how to use technologies to improve their knowledge than previous generations. They utilize identical technologies as other age groups, such as mobile phones. And they rarely utilize learning tools, such as blogs. In addition, they have trouble building knowledge from on-line information, as shown by the incorrect use of search keywords or by a fleeting stay on web pages that precludes an appropriate evaluation. Finally, did you know that most students would prefer to use fewer technological tools in the classroom?

Texto 6

Existe la creencia popular de que las nuevas generaciones de estudiantes poseen habilidades tecnológicas sofisticadas para construir nuevos aprendizajes a partir de información de la web. No obstante, esta creencia es incorrecta. **Surgió cuando en 2001 un asesor y diseñador de videojuegos acuñó el término "nativo digital" para describir a los jóvenes de hoy en día basándose en anécdotas y observaciones informales.** Tanto los indicadores objetivos como la información recogida mediante entrevistas estructuradas a jóvenes de la "nueva era digital" muestran que éstos no tienen un conocimiento mayor de cómo utilizar las tecnologías para mejorar su aprendizaje. Usan tecnologías idénticas a las que usan el resto de grupos, como los móviles, y apenas emplean herramientas de creación de aprendizaje, como los blogs. Además, tienen dificultad para construir conocimiento a partir de información on-line, como muestra el empleo incorrecto de las palabras clave de búsqueda o una permanencia fugaz en las páginas web que impide su correcta evaluación. Por último, ¿sabía que la mayoría de los estudiantes universitarios preferiría que no se usaran tanto las tecnologías en el aula?

Text 7

The belief that listening to the classical music of Mozart increases the intelligence of children is very popular. However, it is now well known to be false. **It arose from a controversial and misrepresented paper published in *Nature* which reported that the spatial intelligence of undergraduate students increased after listening to the classical music of Mozart.** To begin with, the results obtained in assessing the effects of music of Mozart were not conclusive. While some studies obtained improvements, others did not obtain any effect. In addition, in several of these studies, the sample consisted of adults and not of children. Furthermore, in the few cases where improvements were obtained, they occurred in spatial-temporal reasoning, never in general intelligence. Improvements also varied depending on the type of spatial tasks chosen. Moreover when these improvements occurred, they did not last forever but only for a few minutes. Finally, isn't it suspicious that the greatest promoters of Mozart effect are companies and websites of products for children that are making profits of millions of euros?

Texto 7

La creencia de que escuchar música clásica, especialmente de Mozart, aumenta la inteligencia de los niños es muy popular. Sin embargo, hoy en día se sabe que es falsa. **Surgió de un estudio controvertido y tergiversado de la revista Nature donde se encontraba que la inteligencia espacial de un grupo de universitarios mejoraba tras escuchar música de Mozart.** Para empezar, los resultados obtenidos al evaluar los efectos de la música de Mozart no son concluyentes. Mientras algunos estudios científicos encuentran mejoras, otros no. Además, muchos de estos estudios se han hecho con adultos y no con niños. Por otra parte, las pocas investigaciones que han encontrado mejoras, lo hacen únicamente en el razonamiento espacio-temporal, nunca en la inteligencia general. Y además varían en función del tipo de tareas espaciales escogidas. Más aún, estas mejoras, cuando las hay, no se mantienen para siempre sino que apenas duran unos minutos tras escuchar la música. Por último, ¿no es muy sospechoso que los mayores impulsores del efecto Mozart sean empresas y webs de productos infantiles que están haciendo un negocio de millones de euros gracias a él?

Text 8

Several teachers believe that crossed laterality (i.e., right dominant hand and left dominant eye) is associated with learning difficulties. However, this belief is incorrect. **The origin of this belief lies in an intervention method that was created in 1960 to impose "hemispheric dominance" and that has been strongly criticized and censored by the scientific and medical community.** Numerous studies have attempted to verify if there is any association between crossed laterality and academic achievement, but practically none of them has found any result that confirms this link. Most studies clearly show that crossed laterality is not a predictor of lower intelligence in children. Nor is it a sign of more inaccurate or slower reading. Similarly, crossed laterality does not predict spelling difficulties. Nor is it associated with arithmetic difficulties. These results are constant across different age groups. Finally, if crossed laterality is so serious, is it not suspicious that there are no evidence-based interventions to address it yet?

Texto 8

Muchos maestros creen que la lateralidad cruzada (p.ej., mano dominante derecha y ojo dominante izquierdo) es un factor asociado a las dificultades de aprendizaje. Sin embargo, esta idea es incorrecta. **El origen de esta creencia está en un método de intervención que se creó en 1960 para "imponer la dominancia hemisférica" y que ha sido duramente criticado y censurado por la comunidad científica y médica.** Numerosos estudios han intentado comprobar si hay relación entre la lateralidad cruzada y el rendimiento

académico, pero prácticamente ninguno ha encontrado resultados que confirmen esta asociación. La gran mayoría de investigaciones muestra de forma sólida que el hecho de que un niño tenga lateralidad cruzada no es predictor de una menor inteligencia. Tampoco lo es de una lectura más lenta o imprecisa. Asimismo, la lateralidad cruzada no predice las dificultades en escritura. Y tampoco está asociada a dificultades en aritmética. Estos resultados se aplican a niños de cualquier edad. Por último, si tan grave es, ¿no resulta sospechoso que aún no exista ni un solo tratamiento para corregir la lateralidad cruzada con evidencia científica?

Text 9

The idea that the global method (to teach whole words instead of grapheme-phoneme correspondences) is the most appropriate for reading instruction is very popular. However, it is erroneous. **The idea arose when the global method was imported from Anglo-Saxon countries, ignoring that the phonology of English is completely different from ours.** In Spanish, Catalonian, Basque or Galician each sound always corresponds to the same letter, with very few exceptions. If we teach a child these correspondences, they will be able to read independently and accurately any word. English irregularities have fostered the use of the global method, but even Anglo-Saxon countries are returning to the use of grapheme-phoneme correspondences. The numerous studies performed over the last 20 years clearly show the superiority of the synthetic method over the global method to acquire a fluent and accurate reading ability, specifically in children with learning difficulties. In addition, cognitive psychologists have found no evidence that the global method improves reading ability. Finally, did you know that learning the name or the sound of letters is one of the most important predictors of reading success in the future?

Texto 9

La idea de que el método global (enseñar a leer palabras completas y no la asociación letra-sonido) es el más adecuado para la enseñanza de la lectura es muy popular. Ahora bien, esta idea es errónea. **Se inició con la importación de esta metodología desde países anglosajones, sin considerar que la fonología del inglés es muy diferente a la nuestra.** En castellano, catalán, euskera o gallego a cada letra le corresponde siempre un mismo sonido, salvo contadas excepciones. Si a un niño le enseñamos estas correspondencias, podrá leer solo y sin errores cualquier palabra. Las irregularidades del inglés han favorecido el uso del método global pero incluso en los países anglosajones se está retornando a la correspondencia letra-sonido. Los numerosos estudios de los últimos 20 años muestran de forma robusta la superioridad del método sintético sobre el método global para adquirir una lectura fluida y precisa, especialmente en niños con dificultades. Además, desde la psicología cognitiva tampoco hay evidencia de que el método global facilite un aprendizaje lector más significativo. Por último, ¿sabía que aprender el nombre de las letras o su sonido es uno de los principales predictores del éxito lector futuro?