frontiers in Education

SYSTEMATIC REVIEW
published: 31 March 2021
doi: 10.3389/feduc.2021.581216

A Meta-Analysis of Ten Learning Techniques

Gregory M. Donoghue and John A. C. Hattie*

Science of Learning Research Centre, Graduate School of Education, University of Melbourne, Melbourne, VIC, Australia

This article outlines a meta-analysis of the 10 learning techniques identified in Dunlosky et al. (2013a), and is based on 242 studies, 1,619 effects, 169,179 unique participants, with an overall mean of 0.56. The most effective techniques are Distributed Practice and Practice Testing and the least effective (but still with relatively high effects) are Underlining and Summarization. A major limitation was that the majority of studies in the meta-analysis were based on surface or factual outcomes, and there is caution needed when applying these findings to deeper and more relational outcomes. Other important moderators included the presence of feedback or not, near or far transfer, and the effects were much greater for lower than higher ability students. It is recommended that more attention be paid to when, under what conditions, each technique can be used, and how they can best be taught.

Keywords: meta-analysis, learning strategies, transfer of learning, learning technique, surface and deep learning

OPEN ACCESS

Edited by:
Grace Skrzypiec,
Flinders University, Australia

Reviewed by:
Melissa Christine Davis,
Edith Cowan University, Australia
Douglas F. Kauffman,
Medical University of the Americas–Nevis, United States

****Correspondence:***
Gregory M. Donoghue
gregory.donoghue@unimelb.edu.au

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *08 July 2020*
Accepted: *08 February 2021*
Published: *31 March 2021*

Citation:
Donoghue GM and Hattie JAC (2021) A Meta-Analysis of Ten Learning Techniques. Front. Educ. 6:581216. doi: 10.3389/feduc.2021.581216

INTRODUCTION

While the purpose of schooling may change over time and differ across jurisdictions, the mechanisms by which human learning occurs arguably are somewhat more universal. Learning techniques actions that learners themselves can take to enhance their learning–have attracted considerable research interest in recent years (Edwards et al., 2014). This is unsurprising given the direct, practical applicability of such research, and its relevance to students, educators and school leaders alike.

A major, thorough, and important review of various learning techniques has created much interest. Dunlosky et al. (2013a) reviewed 10 learning techniques and a feature of their review is their careful analyses of possible moderators to the conclusions about the effectiveness of these learning techniques, such as learning conditions (e.g., study alone or in groups), student characteristics (e.g., age, ability), materials (e.g., simple concepts or problem-based analyses), and criterion tasks (different outcome measures). This article uses this review as a basis for conducting a meta-analysis on these authors' references to add another perspective of the magnitude of the various learning techniques and how they are affected by various moderators.

Dunlosky et al. (2013a), claim to have conducted an exhaustive search of the literature, relied on previous empirical reviews of learning techniques, and applied a robust set of selection criteria before selecting final 10 techniques. These criteria included that the technique could be implemented by students without assistance, there was sufficient empirical evidence to support at least a preliminary assessment of efficacy, and there was robust evidence to identify the generalizability of its benefits across four categories of variables materials, learning conditions, student characteristics and criterion tasks. Indeed the authors' mastery of the literature is most evident throughout the article.

The authors then categorised the 10 techniques into three groups based on whether they considered them having high, medium or low support for their effectiveness in enhancing learning. Categorised as "high" support were Practice Testing (self-testing or taking practice tests on to-be-learned material) and Distributed Practice (implementing a schedule of practice that spreads out study activities over time in contrast to massed or 'crammed' practice). Categorised as "moderate" support were Elaborative Interrogation (generating an explanation of a fact or concept), Self-Explanation (where the student explains how new information is related to already-known information) and Interleaved Practice (implementing a schedule of practice mixing different kinds of problems within a single study session). Finally, categorised as "low" support were Summarization (writing summaries of to-be-learned texts), Highlighting/Underlining (marking potentially important portions of to-be-learned materials whilst reading), Keyword Mnemonic (generating keywords and mental imagery to associate verbal materials), Imagery use (attempting to form mental images of text materials while reading or listening) and Re-Reading (restudying text material again after an initial reading). In an accompanying article, Dunlosky et al. (2013b) claimed that some of these low support techniques (that students use a lot) have "failed to help students of all sorts" (*p*. 20), the benefits can be short lived, they may not be widely applicable, the benefits are relatively limited, and they do not provide "bang for the buck" (*p*. 21).

Practice Testing is one of the two techniques with the highest utility. This must be distinguished from high stakes testing: Practice Testing instead involves any activity where the student practices retrieval of to-be-learned information, reproduces that information in some form, and evaluates the correctness of that reproduction against an accepted 'correct' answer. Any discrepancy between the produced and "correct" information then forms a type of feedback that the learner uses to modify their understanding. Practice tests can include a range of activities that students can conduct on their own, such as completing questions from textbooks or previous exams, or even self-generated flashcards. According to Dunlosky et al. (2013a), such testing helps increase the likelihood that target information can be retrieved from long-term memory and it helps students mentally organize information that supports better retention and test performance. This effect is strong regardless of test form (multiple choice or essay), even when the format of the practice test does not match the format of the criterion test, and it is effective for all ages of student. Practice Testing works well even when it is massed, but is even more effective when it is spaced over time. It does not place high demand on time, is easy to learn to do (but some basic instruction on how to most effectively use practice tests helps), is so much better than unguided restudy, and so much more effective when there is feedback about the practice test outputs (which also enhances confidence in performance).

Many studies have shown that practice spread out over time (spaced) is much more effective than practice over a short time period (massed)–this is what is meant by Distributed Practice. Most students need three to four opportunities to learn something (Nuthall, 2007) but these learning opportunities are more effective if they are distributed over time, rather than delivered in one massed session: that is, spaced practice, not skill and drill, spread out not crammed, and longer inter-study intervals are more effective than shorter. There have been four meta-analyses of Spaced vs. Massed practices involving about 300 studies, with an average effect of 0.60 (Donovan and Radosevich, 1999; Cepeda et al., 2006; Janiszewski et al., 2003; Lee and Genovese 1988). Cepeda et al. (2008) showed that for almost all retention intervals, memory performance increases sharply with the length of the spacing interval. But at a certain spacing interval, optimal test performance is reached, and from that interval onwards, performance declines but only to a limited degree. But they also note that this does not take into account the absolute level of performance, which decreases as the retention interval increases. Further, Spaced Practice is more effective for deeper than surface processing, and for all ages. Rowland (2014) completed a meta-analysis on 61 studies investigating the effect of testing vs. restudy on retention. He found a high effect size (d = 0.50) supporting the testing over restudy, and the effects were greater for recall than for recognition tasks. The educational message is to review previously covered material in subsequent units of work, time tests regularly and not all at the end (which encourages cramming and massed practice), and given that students tend to rate learning higher after massed, educate them as to the benefits of spaced practice and show them those benefits.

Elaborative Interrogation, Self-Explanation, and Interleaved Practice received moderate support. Elaborative Interrogation involves asking "Why" questions ("Why does it make sense that" "Why is this true") and a major purpose is to integrate new information with existing prior knowledge. The effects are higher when elaborations are precise rather than imprecise, when prior knowledge is higher than lower, and when elaborations are self-generated rather than provided. A constraint of the method is that is more applicable to surface than to deep understanding. Self-explanation involves students explaining some aspect of their processing during learning. It works across task domains, across ages, but may require training, and can take some time to implement. Interleaved Practice involves alternating study practice of different kinds of items, problems, and even subject domains rather than blocking study. The claim is that Interleaving leads to better discrimination of different kinds of problems, more attention to the actual question or problem posed, and as above there is better learning from Spaced than Mass Practice. The research evidence base is currently small, and it is not clear how to break tasks in an optimal manner so as to interleave them.

There is mixed and often low support, claimed Dunlosky et al. (2013a), for Summarization, Highlighting, Keyword Mnemonic, Imagery Use for text learning, and Re-Reading. Summarization involves students writing summaries of to-be-learned texts with the aim of capturing the main points and excluding unimportant or repetitive material. The generality and accuracy of the summary are important moderators, and it is not clear whether it is better to summarize smaller pieces of a text (more frequent Summarization) or to capture more of the text in a larger summary (less frequent Summarization). Younger and less able students are not as good at Summarization, it is better when the assessments are performance or generative and not closed or multiple choice tests, and it can require extensive training to use optimally. Highlighting and Underlining are simple to use, do not require training, and demand hardly any additional time beyond the reading of the text. It is more effective when professionals do the highlighting, then for the student doing the highlighting, and least for reading other student's highlights. It may be detrimental to later ability to make inferences; overall it does little to boost performance. The Keyword Mnemonic involves associating some imagery with the word or concept to be learned. The method requires generating images that can be difficult for younger and less able students, and there is evidence is may not produce durable retention. Similarly Imagery Use is of low utility. This method involves students mentally imaging or drawing pictures of the content using simple and clear mental images. It too is more constrained to imagery-friendly materials, and memory capacity. Re-Reading is very common. It is more effective when the Re-Reading is spaced and not massed, the effects seem to decrease beyond the second reading, is better for factual recall than for developing comprehension, and it is not clear it is effective with students below college age.

A follow-up and more teacher accessible article by Dunlosky et al. (2013b) asks why students do not learn about the best techniques for learning. Perhaps, the authors suggest, it is because curricula are developed to highlight content rather than how to effectively acquire it; and it may be because many recent textbooks used in teacher education courses fail to adequately cover the most effective techniques or how to teach students to use them. They noted that employing the best techniques will only be effective if students are motivated to use them correctly but teaching students to guide their learning of content using effective techniques will allow them to successfully learn throughout their lifetime. Some of the authors' tips include: give a low-stakes quiz at the beginning of each class and focus on the most important material; give a cumulative exam that encourages students to re-study the most important material in a distributed fashion; encourage students to develop a "study planner" so they can distribute their study throughout a class and rely less on cramming; encourage students to use practice retrieval when studying instead of passively re-reading their books and notes; encourage students to elaborate on what they are reading, such as by asking "why" questions; mix up problems from earlier classes so students can practice identifying problems and their solutions; and tell students that highlighting is fine but only in the beginning of their learning journey.

The Dunlosky et al. (2013a), review shows a high level of care of selection of articles, an expansiveness of the review, an attention to generalizability and moderators, and is sophisticated in its conclusions. There are two aspects of the this research that the current paper aims to address. First, Dunlosky et al. (2013a) relied on a traditional literature review method and did not include any estimates of the effect-sizes of their various techniques, nor did they indicate the magnitude of their terms high, medium, and low. One of the purposes of this article is to provide these empirical estimates. Second, the authors did not empirically evaluate the moderators of the 10 learning techniques, such as Deep vs. Surface learning, Far vs. Near Transfer, or age/grade level of learner. An aim of this paper is to analyze the effects of each of the 10 techniques with respect to these and other potential moderators.

METHOD

Research syntheses aim to summarise past research by estimating effect-sizes from multiple, separate studies that address, in this case, 10 major learning techniques. The data is based on the 399 studies referenced in Dunlosky et al. (2013a). We removed all non-empirical studies, and any studies that did not report sufficient data for the calculation of a Cohen's *d*. This resulted in 242 studies being included in the meta-analysis, many of which contained data for multiple effect sizes, resulting in 1,620 cases for which a Cohen's *d* was calculated (see **Figure 1**).

The publication dates of the articles ranged from 1929 to 2014, with half being published since 1996. Most participants were undergraduates (65%), but also included secondary (11%), primary (13%), adults (2%), and early childhood (9%). Most were chosen from the average range of abilities (86%), while 7% were categorised low ability and 7% high ability. The participants were mainly North Americans (86%), and also included Europeans (11%), and Australians (3%).

All articles were coded by the two authors, and independent colleagues were asked to re-code a sample of 30 (about 10%) to estimate inter-rater reliability. This resulted in a kappa value of 0 89, which gives much confidence in the dependability of the coding.

For each study, three sets of moderators were coded. The first set of moderators included attributes of the article: quality of the journal (h-index), year of publication (to assess any changes in effectiveness as more research has been added into the literature), and sample size. The second set of moderators included attributes of the students: ability level of the students (low, average, and high), country of the study, grade levels of the student (pre and primary, high, Univ, adults). The third set of moderators included attributes of the design: whether the outcome was near or far transfer (e.g., was the learner tested on criterion tasks that differed from the training tasks or did the effect of the technique improve the student learning in a different subject domain), the depth of the outcome (Surface or content-specific vs. Deep or more generalizable learning), how delayed was the testing from the actual study (under 1 day, or 2 + days), the learning domain of the content of the study or measure (e.g., cognitive, non-cognitive).

The design of most studies include experimental compared to control group (91%), longitudinal (pre-post, time series) 6.5%, and within subject designs (2.4%). Most learning outcomes were classified as Surface (93%) and the other 7% Deep. The post-tests were predominantly completed very soon after the intervention - 74% completed in 1 day or less, 17% from 2 to 7 days, 3.3% from 8 days to month, 0.4% from 1 to 3 months, and 0.2% from 4 months to 7 years.

We used two major methods for calculating Cohen's *d* from the various statistics published in the studies. First, standardized mean differences (N = 1,203 effects) involved subtracting the mean of the control group from the mean of the experimental

TABLE 1 | Summary of effects for each learning strategy.

Learning strategy	Dunlosky classification	# Cases	Unique N	d	SEM	q	I^2 (%)
Distributed practice	High	150	152,952	0.85	0.053	887.0	83
Practice testing	High	374	6,033	0.74	0.04	2,613.3	86
Elaborative interrogation	Moderate	254	2,138	0.56	0.048	1,172.4	78
Imagery	Moderate	135	1,052	0.56	0.061	415.9	68
Self explanation	Moderate	93	804	0.54	0.092	394.0	77
Mnemonics	Low	107	580	0.50	0.104	933.9	89
Re-reading	Low	113	1,529	0.47	0.06	792.3	86
Interleaved practice	Low	104	972	0.47	0.089	864.3	88
Underlining	Low	56	1,129	0.44	0.115	242.0	77
Summarization	Low	234	1,990	0.44	0.055	2063.4	89
Average/Total		1,619	169,179	0.56		10,688	85

group, then dividing by an estimate of the pooled standard deviation, as follows.-

$$\text{Cohen's } d = \frac{\overline{x}_e - \overline{x}_c}{\mathrm{SD_{pooled}}} \quad \text{where } \mathrm{SD_{pooled}} = \frac{\mathrm{SD}_e + \mathrm{SD}_c}{2}$$

The standard errors of the effect size (ES) were calculated as follows,

$$\mathrm{SE} = \sqrt{\left(n_1 + \frac{n_1}{n_1 * n_1}\right) + \left[\frac{\mathrm{ES} * \mathrm{ES}}{2\,(n_c + n_e)}\right]}$$

We adjusted the effect sizes (ES) according to Hedges and Olkin, (1985) to account for bias in sample sizes, according to this

$$\mathrm{ESg} = \mathrm{ES}^{*}\left\{1 - \frac{3}{(4\mathrm{N} - 9)}\right\}$$

Second, F-statistics (for two groups only) were converted using:

$$\mathrm{ESf} = \sqrt{F * \frac{\{n_c + n_e\}}{\{n_c * n_e\}}}$$

The Standard Error was calculated using:

$$\mathrm{SE} = \frac{ES}{\sqrt{ES * N}}$$

In all cases, therefore, a positive effect meant that the learning technique had a positive impact on learning.

The distribution of effect sizes and sample sizes was examined to determine if any were statistical outliers. Grubbs (1950) test was applied (see also Barnett and Lewis, 1994). If outliers were identified, these values were set at the value of their next nearest neighbour. We used inverse variance weighted procedures to calculate average effect sizes across all comparisons (Borenstein et al., 2005). Also, 95% confidence intervals were calculated for average effects. Possible moderators (e.g., grade level, duration of the treatment) of the DBP to student outcome relationship were tested using homogeneity analyses (Hedges and Olkin, 1985; Cooper et al., 2019). The analyses were carried out to determine whether a) the variance in a group of individual effect sizes varies more than predicted by sampling error and/or b) multiple groups of average effect sizes vary more than predicted by sampling error.

Rather than opt for a single model of error, we conducted the overall analyses twice, once employing fixed-error assumptions and once employing random-error assumptions (see Hedges and Vevea, 1998, for a discussion of fixed and random effects). This sensitivity analysis allowed us to examine the effects of the different assumptions (fixed or random) on the findings. If, for example, a moderator is found to be significant under a random-effects assumption but not significant under a fixed effects assumption, then this suggests a limit on the generalizability of the inferences of the moderator. All statistical processes were conducted using the Comprehensive Meta-Analysis software package (Borenstein et al., 2005).

The examination of heterogeneity of the effect size distributions within each outcome category was conducted using the Q statistic and the I^2 statistic (Borenstein et al.,

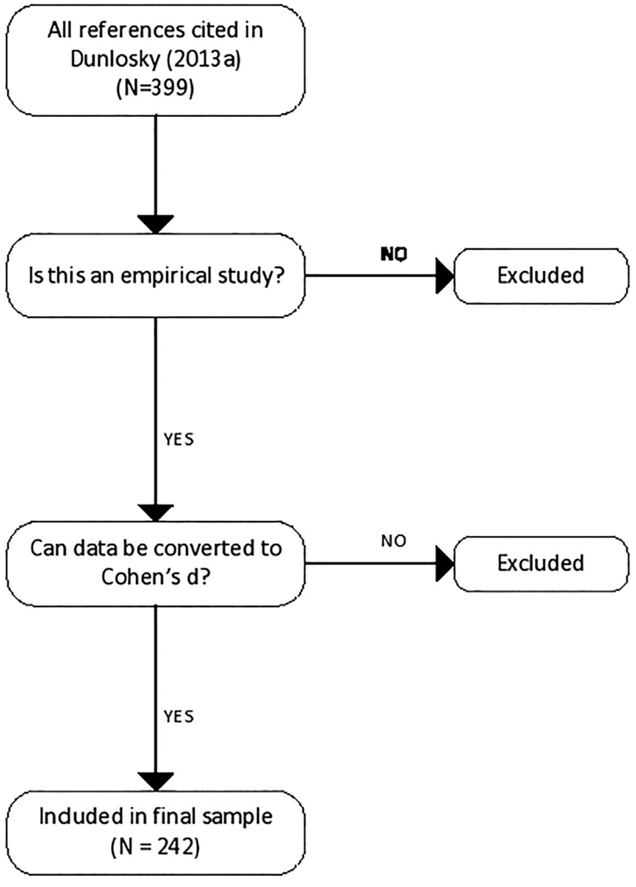
All references cited in Dunlosky (2013a) (N=399)
Is this an empirical study?
NO
Excluded
YES
Can data be converted to Cohen's d?
NO
Excluded
YES
Included in final sample (N = 242)

FIGURE 1 | Flow diagram of articles used in the meta-analysis.

TABLE 2 | Effect Sizes moderated by the Learning Domain.

Effect sizes	English	General knowledge	Humanities	Languages	Mathematics	Recall	Science	Unknown
Elaborative interrogation	0.38		–0.11			0.46	0.76	
Self explanation	0.36				0.50		0.63	
Summarization	0.39		–0.08			0.59	0.77	
Underlining	0.42		–0.19	2.16		0.52	0.50	
Mnemonics	1.17			0.75		–0.34		
Imagery	0.51			2.98		1.01	0.29	
Re-reading	0.54		0.44				–0.04	
Practice testing	0.94	0.52	0.70	1.29	0.13	0.71	0.64	
Distributed practice	0.88		0.88	0.67	0.39	1.16	0.63	0.61
Interleaved practice	0.10		0.99		1.66	0.31	0.20	
Average	0.57	0.52	0.38	1.57	0.67	0.55	0.49	0.61

TABLE 3 | Effect sizes moderated by grade level.

S	Prim			Sec			Univ		
	Mean	SE	N	Mean	SE	N	Mean	SE	N
Distributed practice	0.57	0.11	15	0.70	0.24	5	0.89	0.06	115
Elab interrogation	0.38	0.14	28	0.71	0.06	84	0.58	0.09	110
Imagery	0.38	0.06	101	0.82	0.17	8	1.16	0.18	26
Interleaved practice	–0.20	0.15	13	0.00	0.07	12	0.65	0.11	79
Mnemonics	0.00	0.00	6	0.69	0.17	29	0.20	0.13	72
Practice testing	–0.25	0.56	6	0.99	0.24	22	0.80	0.04	278
Re-reading	1.33	0.26	6				0.42	0.06	107
Self explanation	0.28	0.09	15	0.57	0.16	36	0.60	0.15	42
Summarization	1.20	0.13	25	0.75	0.14	58	0.19	0.05	151
Underlining	–0.06	0.18	10	–0.10		1	0.57	0.14	43
Total	0.42	0.05	225	0.69	0.05	255	0.59	0.03	1,023

2009). To calculate Q and I^2, we entered the corrected effect-sizes for every case, along with the SE (calculated as above) and generated homogeneity data. Due to the substantive variability within the studies, even in the case of a non-significant Q test, when I^2 was different from zero, moderation analyses were carried out through subgroup analysis (Lipsey and Wilson, 2001). As all hypothesized moderators were operationalized as categorical variables, these analyses were performed primarily through subgroup analyses using a mixed-effects model.

RESULTS

Table 1 shows a comprehensive analysis of the collected data. For the 242 studies, we calculated a total of 1,619 effects which related to 169,179 unique participants. The overall mean assuming a fixed model was 0.56 (SD = 0.81, SEM = 0.072, skewness 1.3, kurtosis = 5.64); the overall mean assuming a random model was 0.56 (SE = 0.016). The overall mean at the study level was 0.77 (SE = 0.049). The fixed effects model assumes that all studies in the meta-analysis share a common true effect size, whereas the random effects model assumes that the studies were drawn from populations that differ from each other in ways that could impact on the treatment effect. Given that the means estimated under the two models are similar we proceed to use only one (the random model) in subsequent analyses.

The distribution of all effects is presented in **Figure 1** and the studies, their attributes, and study effect-size are presented in **Table 1**. It is clear that there is much variance among these effects (Q = 10,688.2, I^2 = 84.87). The I^2 a measure of the degree of inconsistency in the studies' results; and this I^2 of 85% shows that most of the variability across studies is due to heterogeneity rather than chance. Thus, the search for moderators is critical to understanding which learning techniques work best in which situations.

Table 2 shows an overall summary of effects moderated by the learning domain. The effects correspond with the classification of High, Moderate, and Low by Dunlosky et al. (2013a), but it is noted that Low is still relatively above the average of most meta-analysis in education – Hattie, (2009), Hattie, (2012), Hattie, (2015) reported an average effect-size of 0.40 from over 1,200 meta-analyses relating to achievement outcomes. All techniques analyzed in the current study had an ES of over 0.40.

Moderator Analyses

Year of publication

There was no relation between the magnitude of the effects and the year of the study (r = 0.08, df = 236, p = 0.25) indicating that the effects of the learning technique have not changed over time (from 1929 to 2015).

TABLE 4 | Effect size moderated by Country of first author.

	Australia	Canada	Europe	USA	Total
Distributed pract ice			0.56	0.88	0.86
Elab interrogation		0.76		0.27	0.56
Imagery			0.88	0.52	0.56
Interleaved practice			0.21	0.62	0.47
Mnemonics			−0.44	0.56	0.33
Practice testing	0.46		0.67	0.78	0.76
Re-reading				0.47	0.47
Self explanation	0.72		0.57	0.38	0.54
Summarization			0.29	0.44	0.44

Learning Domain

The vast majority of the evidence is based on measurements of academic achievement: 222 of the 242 studies (91.7%) and 1,527 of the 1,619 effects (94.3%). English or Reading was the basis for 85 of the studies (35.1%) and 546 of the effects (33.7%), and Science 41 of the studies (16.9%) and 336 of the effects (20.8%). There was considerable variation in the effect sizes of these domains, as shown in **Table 3**.

Near vs. Far Transfer

If the study measured an effect on performance on a task similar to the task used in the experiment, it was classified as measuring Near transfer, alternatively if the transfer was to another dissimilar context it was classified as Far transfer. There were so few Far transfer effects that the information is not broken into the 10 learning techniques. Overall, the effects on Near ($d = 0.61$, SE = 0.052, $N = 197$) are much greater than the effects on Far ($d = 0.39$, SE = -0.002, $N = 1,385$).

Depth of Learning

The effects were higher for Surface ($d = 0.60$, SE = 0.021, $N = 1,473$) than for Deep processing ($d = 0.26$, SE = 0.064, $N = 109$).

Grade Level

The effects moderated by grade level of the participants are presented in **Table 4**. All students had higher effects on summarization, distributed practice, imagery use, and re-reading, primary students had lower effects on interleaved practice, mnemonics, self-explanation, and practice testing. Both primary and secondary students had lower effects on Underlining.

Country

Each study was coded for the country where the study was conducted. Where that information was not made clear in the article, the first author's country of employment was used. Of the 242 studies, 187 (77.3%) were from USA, 20 (8.3%) were from Canada, 27 (11.1%) from Europe: United Kingdom, Denmark, France, Germany, Italy, Netherlands), 7 (2.9%) from Australia and 1 (0.4%) from Iran making a total North American proportion of 207 (85.6%). Other than the drop for Europe in Mnemonics, Interleaved Practice and Summarisation there is not a great difference by country.

Ability Level

Almost all studies referred to participants as being either "Low" "Normal" or "High" ability. This language has been continued in the collection and analysis of the data, however in the body of the paper the terms "Low", "Average" and "High" ability have been used instead. In all cases, these categories aligned with percentiles of the normal distribution for academic scores. Of the 242 studies, only six investigated High ability students, and only 13 Low ability. Across all techniques, the mean effect on High ability students was -0.11 (SE = 0.10, $N = 28$) for Low ability students was 0.47, SE = 0.15, $N = 58$. The High ability students had negative effects for Interleaved Practice and Summarization.

Delay

Studies predominantly measured only very short-term effects, the exception being the three learning techniques focused on practice effects (Practice Testing, Distributed Practice and Interleaved Practice). Most (68%) where evaluated within a day (usually immediately). There were no overall differences relating to the effects less than a day (d = 0.58, SE = 0.025, $N = 1,073$), > 1 day and < 1 week ($d = 0.59$, SE = 0.057, $N = 204$), > 1 week and < 1 month ($d = 0.56$, SE = 0.058, $N = 228$), < 1 month and less than 6 months ($d = 0.51$, SE = 0.082, $N = 64$).

Journal Impact Factor

The published Impact factor for each journal was sourced from that Journal's website. Where a multiple-year (usually 5 years) average h-index was provided, that was used in preference to a single (the most recent) year (PhD theses were left blank). The average impact factor is 2.80 (SE = 3.29), which relative to Journals in educational psychology indicates that the overall quality of Journals is quite high. Across all 10 learning techniques, there was a moderate positive correlation between effect size and Journal Impact Factor, r (235) = 0.24, < 0.001. Thus the effect-sizes were slightly higher in the more highly cited Journals.

DISCUSSION AND CONCLUSION

The purpose of the current study was twofold: to provide empirical estimates of the effectiveness of the 10 learning techniques, and second, to empirically evaluate a range of their potential moderators. The major conclusion from the meta-analysis is a confirmation of the major findings in Dunlosky et al. (2013a). They rated the effects by High, Moderate, and Low and there was much correspondence between their ratings and the actual effect-sizes: High in the meta-analysis was > 0.70, Moderate between 0.54 and 0.69, and Low < 0.53. This meta-analysis, however, shows the arbitrariness of these ratings, as some of the low effects were very close estimates to the moderate. mnemonics, re-reading and interleaved practice were all within 0.06 of the moderate category and these techniques may have similar importance to those Dunlosky et al. (2013a) classified as Moderate. Certainly they should not be dismissed as ineffective. Even the lowest

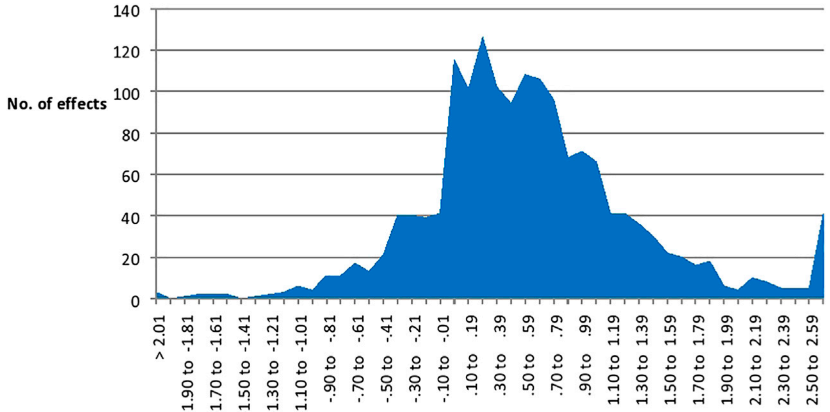

FIGURE 2 | Distribution of effects.

learning techniques (Underlining and Summarization (both *d* = 0.44) are sufficiently effective to be included in a student's toolbox of techniques.

The rating method into High, Medium, and Low was matched by the findings of the meta-analysis, but **Table 2** shows the usual difficulties of such arbitrary (but not capricious) cut scores. Mnemonics (*d* = 0.50) is close to Self-Explanation (*d* = 0.54), although there is a clear separation between Moderate (Elaborative Interrogation *d* = 0.56) and Practice Testing (*d* = 0.74). All have a sufficient positive effect to consider by students choosing learning techniques, and it may be that there is a more optimal stage during the learning process to choose the high techniques related to consolidating learning, and the low techniques related to first encountering new material and ideas. It may also be that techniques are affected by whether the tasks are more relevant to memory vs. those that are relevant to comprehension. Many of the techniques in the authors' list of 10 are more related to the former than the latter.

The technique with the lowest overall effect was Summarization. Dunlosky et al. (2013a) note that it is difficult to draw general conclusions about its efficacy, it is likely a family of techniques, and should not be confused with mere copying. They noted that it is easy to learn and use, training typically helps improve the effect (but such training may need to be extensive), but suggest other techniques might better serve the students. In their other article (Dunlosky et al., 2013b), the authors classified Summarization as among the "less useful techniques" that "have not fared so well when considered with an eye toward effectiveness" (*p*. 19). They also noted that a critical moderator for the effectiveness of all techniques is the student's motivation to use them correctly. This meta-analysis shows that Summarization, while among the less effective of the 10 under review, still has a sufficiently high impact to be considered worthwhile in the student's arsenal of learning techniques, and with training could be among the more easier to use techniques.

One of the sobering aspects of this meta-analysis is the finding that the majority of studies are based on Surface learning of factual, academic content, measure learning almost immediately after the technique has been used, and only measure Near transfer. This limits the generalisability of the Dunlosky et al. (2013a) review and this meta-analysis and there may well be different learning techniques that optimise deeper learning, non-academic learning, or more intensive learning that requires longer retention periods and Far transfer. The verdict is still out on the effectiveness and identification of the optimal techniques in these latter conditions. It should be noted, however, that this may be not only a criticism of the current research on learning techniques but could well be the same criticism of student experiences in most classrooms. Too many modern classrooms are still dominated by a preponderance of surface learning, teachers asking low level questions demanding content answers, and assessments privileging surface knowledge (Tyack & Cuban, 1995). Thus the 10 techniques may remain optimal for many current classrooms.

The implication for teachers is not that these learning techniques should be implemented as stand-alone "learning interventions" or fostered through study skills courses. They can be used, however, within a teaching process to maximise the surface and deeper outcomes of a series of lessons. For example, Practice Testing is among the top two techniques but it would be a mistake to then make claims that there should be more testing, especially high-stakes testing! Dunlosky et al. (2013a) concluded that more Practice Testing is better, should be spaced not massed, works with all ages, levels of ability, and across all levels of cognitive complexity. A major moderator is whether the practice tests are accompanied by feedback or not. "The advantage of Practice Testing with feedback over restudy is extremely robust. Practice Testing with feedback also consistently outperforms Practice Testing alone" (*p*. 35). If students continue to practice wrong answers, errors or misconceptions, then these will be successfully learnt and become high-confidence errors; hence the power of feedback. It is not the frequency of testing that matters, but the skill in using practice testing to learn and consolidate knowledge and ideas.

There are still many unanswered questions that need further attention. First, there is a need to develop a more overarching model of learning techniques to situate these 10 and the many other learning techniques. For example, we have developed a model that argues that various learning techniques can be optimised at certain phases of learning from Surface to Deep to Transfer, from acquiring and consolidating knowledge and understanding, and involves three inputs and outputs -knowing, dispositions, and motivations; which we call the skill, the will, and the thrill (Hattie and Donoghue, 2016). Memorisation and Practice Testing, for example, can be shown to be effective in the consolidating of surface knowing but not effective without first acquiring surface knowing. Problem based learning is relatively ineffective for promoting surface but more effective at the deeper understanding, and thus should be optimal after it has been shown students have sufficient surface knowledge to then work through problem based methods.

Second, it was noted above that the preponderance of current studies (and perhaps classrooms) favour Surface and Near learning and care should be taken to not generalise the results of either the original review or our meta-analysis to when Deep and Far learning is desired. Third, it is likely, as the original authors hint, having a toolbox of optimal learning techniques may be most effective, but we suggest that there may need to be a higher sense of self-regulation to know when to use them. Fourth, as the authors noted, it is likely that motivation and emotions are involved in the selection, persistence with, and effectiveness of using the learning techniques, so attention to these matters is imperative for many students. Fifth, given the extensive and robust evidence for the efficacy of these learning techniques, an important avenue of future research may centre on the value in teaching them to both teachers and students. Can these techniques be taught, and if so, how? Need they be taught in the context of specific content? In what ways can the emerging field of educational neuroscience inform these questions?

Third, Dunlosky and Rawson (2015) noted that more recent research may influence some of these findings. For example, he noted that while Interleaving was a “Low” technique, there have since been many studies demonstrating the benefits of Interleaving. For example, Carvalho and Goldstone (2015) found that the way information is ordered impacts learning and that this influence is modulated by the demands of the study task; in particular whether learning is active or passive. Learners in the active study condition tend to look for features that discriminate between categories, and these features are easier to detect when categories frequently alternate (i.e., using Interleaving). Learners in the passive study condition are more likely to look for features that consistently appear within one category’s examples, and these features are easier to detect when categories rarely alternate.

Limitation

A significant limitation of the current study is that no publications beyond 2014 have been meta-analysed. Notwithstanding, the authors are unaware of any more recent study that contradicts any of our findings. Accordingly, the study represents a comprehensive and valid quantitative review of research published between 1929 and 2014, one that complements and underpins Dunlosky et al. (2013a) qualitative review.

CONCLUDING REMARKS

The major contribution from Dunlosky et al. (2013a), and supported by the findings from this study is to highlight the relative importance of learning techniques, to identify and allow for the optimal moderators, and clearly more defensible models are needed that take into account the demands of the task, the timing of the intervention, and the role of learning techniques within content domains. Future research that examines the impact of these (and other) moderators, and incorporates findings into theoretical and conceptual models, is much needed.

AUTHOR CONTRIBUTIONS

JH conceived study, wrote article with GD. GD found and coded all article, worked on analyses, contributed to writing.

REFERENCES

Barnett, V., and Lewis, T. (1994). *Outliers in statistical data*. New York, NY: Wiley.

Borenstein, M., Cooper, H., Hedges, L., and Valentine, J. (2009). Effect sizes for continuous data. *Handbook Res. Synth. Meta-Anal.* 2, 221–235. doi:10.7758/9781610448864.4

Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2005). *Comprehensive meta-analysis version 2.* Englewood, NJ: Biostat.

Carvalho, P. F., and Goldstone, R. L. (2015). The benefits of interleaved and blocked study: different tasks benefit from different schedules of study. *Psychon. Bull. Rev.* 22 (1), 281–288. doi:10.3758/s13423-014-0676-4

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132 (3), 354. doi:10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., and Pashler, H. (2008). Spacing effects in learning: a temporal ridgeline of optimal retention. *Psychol. Sci.* 19 (11), 1095–1102. doi:10.1111/j.1467-9280.2008.02209.x

H. Cooper, L. V. Hedges, and J. C. Valentine (Editors) (2019). *The handbook of research synthesis and meta-analysis* (Newyork, NY: Russell Sage Foundation).

Donovan, J. J., and Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: now you see it, now you don’t. *J. Appl. Psychol.* 84 (5), 795. doi:10.1037/0021-9010.84.5.795

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., and Willingham, D. T. (2013a). Improving students’ learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychol. Sci. Public Interest* 14 (1), 4–58. doi:10.1177/1529100612453266

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., and Willingham, D. T. (2013b). What works, what doesn’t. *Sci. Am. Mind* 24 (4), 46–53. doi:10.1038/scientificamericanmind0913-46

Dunlosky, J., and Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: tips for classroom use and for guiding students’ learning. *Scholarship Teach. Learn. Psychol.* 1 (1), 72. doi:10.1037/stl0000024

Edwards, A. J., Weinstein, C. E., Goetz, E. T., and Alexander, P. A. (2014). *Learning and study techniques: issues in assessment, instruction, and evaluation*. Amsterdam, The Netherland: Elsevier.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Statist.* 21 (1), 27–58. doi:10.1214/aoms/1177729885

Hattie, J. A., and Donoghue, G. M. (2016). Learning techniques: a synthesis and conceptual model. *Npj Sci. Learn.* 1, 16013. doi:10.1038/npjscilearn.2016.13

Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship Teach. Learn. Psychol.* 1 (1), 79. doi:10.1037/stl0000021

Hattie, J. (2012). *Visible learning for teachers: maximizing impact on learning*. England, United Kingdom: Routledge.

Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. England, United Kingdom: Routledge.

Hedges, L. V., and Olkin, I. (1985). *Statistical methods for meta-analysis*. Cambridge, MA: Academic Press.

Hedges, L. V., and Vevea, J. L. (1998). Fixed-and randomeffects models in meta-analysis. *Psychol. Meth.* 3, 486.

Janiszewski, C., Noel, H., and Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: implications for research on advertising repetition and consumer memory. *J. Consum. Res.* 30 (1), 138–149. doi:10.1086/374692

Lee, T. D., and Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Res. Q. Exerc. Sport* 59 (4), 277–287. doi:10.1080/02701367.1988.10609373

Lipsey, M. W., and Wilson, D. B. (2001). Practical meta-analysis. Newbury Park, CA, United States: SAGE publications, Inc.

Nuthall, G. (2007). *The hidden lives of learners*. Wellington, New Zealand: NZCER Press.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol. Bull.* 140 (6), 1432. doi:10.1037/a0037559

Tyack, D. B., and Cuban, L. (1995). *Tinkering toward utopia*. Cambridge, MA: Harvard University Press.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.