



# Effect of active learning versus traditional lecturing on the learning achievement of college students in humanities and social sciences: a meta-analysis

Anastassis Kozanitis<sup>1</sup> · Lucian Nenciovici<sup>1</sup>

Accepted: 25 November 2022 / Published online: 7 December 2022  
© The Author(s) 2022

## Abstract

A previous meta-analysis found that active learning has a positive impact on learning achievements for college students in STEM fields of study. However, no similar meta-analyses have been conducted in the humanities and social sciences. Because major dissimilarities may exist between different fields or domain of knowledge, there can be issues with transferring research findings or knowledge across fields. We therefore meta-analyzed 104 studies that used assessment scores to compare the learning achieved by college students in humanities and social science programs under active instruction versus traditional lecturing. Student performance on assessment scores was found to be higher by 0.489 standard deviations under active instruction ( $Z=6.521$ ,  $p<0.001$ ,  $k=111$ ,  $N=15,896$ ). The relative beneficial effect of active instruction was found to be higher for some course subject matters (i.e., Sociology, Psychology, Language, Education, and Economics), for smaller ( $\leq 20$  students) rather than larger class or group sizes, and for upper level rather than introductory courses. Analyses further suggest that these findings are not affected by publication bias.

**Keywords** Meta-analysis · Active learning; Humanities · Social sciences · Learning achievement · Higher education

In recent years, researchers in the field of higher education have become increasingly interested in assessing traditional instruction practices (e.g., lectures) and modifying them towards more student-centered and active instructional approaches. As such, multiple political organizations (e.g., UNESCO), professional associations (e.g., European Society for Engineering Education), and accrediting organizations (e.g., Accreditation Board for Engineering and Technology) have recommended the use of more active instructional methods in higher education (Grosemans et al., 2017; Hartikainen et al., 2019; Lima et al., 2017). One important reason identified in previous research that may explain this shift is that student-centered and active instructional methods lead to greater achievement from the viewpoint of student learning outcomes when compared to traditional, more content-centered,

---

✉ Anastassis Kozanitis  
kozanitis.anastassis@uqam.ca

<sup>1</sup> Department of Didactics, Université du Québec À Montréal, Sainte-Catherine East, Montreal, QC 405H2L 2C4, Canada

and passive approaches, such as lecturing (Burgess et al., 2014; Hofer et al., 2018; Swanson, et al., 2017). Abundant evidence is widely available in the scientific literature, especially through access to meta-analysis. For instance, the largest, most comprehensive, and most frequently cited research on the superiority of active instruction for learning achievements in the context of higher education, as compared to passive instruction, is the meta-analysis conducted by Freeman et al. (2014). They meta-analyzed 158 studies that reported data on examination scores comparing student performance in undergraduate science, technology, engineering, and mathematics (STEM) courses under both active learning methods and traditional lecturing. The calculated effect sizes indicated that student performance on examination scores was on average 0.47 standard deviations higher under active learning and that this result was statistically significant ( $g=0.47$ ,  $Z=9.781$ ,  $p<0.001$ ).

Although comprehensive, one important limitation of Freeman et al.'s meta-analysis is that it focused only on studies covering the fields of study in science, technology, engineering, and mathematics (STEM), excluding the fields of study in the humanities and social sciences. A field of study is a broad demarcation known as an area or a branch of knowledge that shares similar theoretical bases. It is taught as an accredited part of higher education, where a university degree is awarded upon completion of a certain number of required and/or elective courses. A course usually provides knowledge pertaining to a field of study. According to the authors who have examined issues with transferring research findings or knowledge across fields of study, there are several reasons why it would be difficult or impossible to simply transfer Freeman et al.'s findings to the fields of study in the humanities and social science areas (Adler et al., 2018; Bartha, 2013; Brunner, 2010; Cartwright & Hardie, 2012). This type of transfer requires a thorough examination of the similarities and differences between the *context* in which the original findings or knowledge was produced and the one to which they are to be transferred. For instance, some of these differences can be related to the nature of the knowledge, the skills, and the learning outcomes encountered in each specific field of study (Biggs, 2003; Hill & Jordan, 2021). These differences can also be related to students' background knowledge and preferred learning style (John et al., 2016), to instructor's posture and perception of his role as a teacher (Deschryver & Lameul, 2016; Pratt, 1992), and to expected student–teacher relationship within a field of study (Tormey, 2021). For these reasons, it would appear inadvisable to simply transfer Freeman et al.'s findings into the field of humanities and social sciences for the purpose of informing higher education policymakers and practitioners. As such, a new meta-analysis in the field of humanities and social sciences, equivalent to the work by Freeman et al., would seem necessary, although we can expect to find similar outcomes based on previous literature reviews.

To our knowledge, no meta-analysis equivalent to Freeman et al.'s work has yet been published to comprehensively quantify the effect of active learning in higher education for the humanities and social sciences. The secondary literature published on this field of research covers it in a non-comprehensive manner and appears to mostly consist of narrative, non-quantitative literature reviews focused on instructional strategies for active learning as a whole (e.g., Michael, 2006; Patton, 2015; Prince, 2004) or on specific strategies like flipped classrooms (e.g., Bishop & Verleger, 2012; Uzunboylu & Karagozlu, 2015), classroom response systems (e.g., Fies & Marshall, 2006; Rana et al., 2016; Simpson & Oliver, 2006), or project-based learning (e.g., Condliffe et al., 2017; Kokotsaki et al., 2016; Thomas, 2000). Moreover, published meta-analyses on specific active instruction methods, such as problem-based learning (e.g., Dochy et al., 2003; Newman, 2003), flipped classrooms (e.g., Rahman et al., 2014), and team-based learning (e.g., Swanson et al., 2017), included very few, if any, primary studies in the fields of humanities and social sciences

and drew no conclusions as to the effectiveness of these active interventions in those fields specifically. The general objective of the present work is to conduct a meta-analysis equivalent to Freeman et al. (2014) that exclusively considers studies in the fields of humanities and social sciences. We hope it will provide instructors in these fields with sound scientific research-based evidence for them to choose their teaching methods.

## Conceptual framework

Active learning is a broad concept. Research into the topic often leaves it vaguely or only implicitly defined, or does not define it at all. When it is explicitly defined, a variety of perspectives can be found among the definitions available in the literature (Drew & Mackie, 2011; Hartikainen, et al., 2019; Prince, 2004). In their often-cited book, Watkins, Lodge, and Carnell (2007) recognized this variety of perspectives and proposed a tripartite framework to help analyze the various definitions in use. The three aspects of active learning under this framework are *behavioral* (e.g., actively using and developing resources), *cognitive* (e.g., actively thinking about experiences to make sense of them and promote knowledge construction), and *social* (actively interacting with others as both collaborators and resources). In that sense, activating instructional methods are student-centered and can therefore include any teaching method that gets students to do something, either behaviorally, cognitively, or socially, for example, questioning, discussing, writing, problem-solving, doing any kind of teamwork, peer learning, and being involved in hands-on experiments. We have opted for this definition of active learning. These activating methods are generally contrasted with traditional instruction methods, such as the lecture or teacher presentations (Bonwell & Sutherland, 1996; McKeachie, 2011). Such methods are considered instructor-centered because they put the burden of communicating course material on the instructor, and have students mostly involved in passive listening. Also, although there is no consensus or generally agreed upon categorization of active learning teaching methods, many authors use broad characteristics allowing for such categorization: collaborative learning, cooperative learning, inquiry-based learning, or experiential learning (Bonwell & Eison, 1991; Kozanitis & Quévillon-Lacasse, 2018; Prosser & Trigwell, 2013). Collaborative learning and cooperative learning are very similar; in both cases, students are asked to interact in small teams in order to complete a task. Examples of such teaching methods include jigsaw activity, small group discussions, and think-pair-share. Teaching methods that are considered inquiry-based learning encourage students to research topics, to develop questions, and to explore problems and reflect on how to solve them. Problem-based learning and case methods are well-known examples of inquiry-based learning. Experiential learning involves learning through judiciously chosen experiences and supported by critical analysis and reflection. Experiential learning usually emulates real-world environments, where students take upon roles. Project-based learning, simulations, and role playing are eminent examples of experiential learning.

Other representations for categorizing active learning teaching methods can also be found. For example, Lord, Prince, Stefanou, Stolk, and Chen (2012) proposed the active learning continuum, where on one end of the continuum are instructor-centered simple and short active learning activities (e.g., think-pair-share), and on the other end are more elaborate teaching methods such as project-based learning and problem-based learning, which are considered to be more student-centered learning activities. It is important to point out that, although differences do exist between any given active learning teaching

method, an overview of the literature suggests that they all share a common goal, which is to actively engage students in the learning process. For this reason, we are not going to dwell on the different definitions and descriptions of the active learning methods. We encourage, for this purpose, to consult the work of Prince (2004), Barkley (2009), Barkley et al. (2014).

Thus, this study aims specifically to meta-analyze studies by comparing learning achievements under active instruction methods and traditional lecturing for college courses in humanities and social science programs. To facilitate comparison with Freeman et al.'s (2014) study evoked earlier, the same confounding variables, when possible, will be considered for heterogeneity analyses, with respect to examination scores. Therefore, the following null hypothesis will be tested:

$H_{0(1)}$ : Assessment scores for students under active learning will not differ significantly from assessment scores for students under traditional lecturing.

Moreover, subgroup analyses will be conducted to test whether there is any significant variance in examination scores using the same study characteristics as Freeman et al. In brief, Freeman et al.'s study considered the following characteristics: courses in the *STEM fields of study* (e.g., biology, physics, engineering, computer science), *assessment type* (concept inventory [standardized test], instructor-written course exams), *class size* (small [ $\leq 50$ ], medium [ $50 < n \leq 110$ ], large [ $> 110$ ]), *course level* (introductory, upper division), and *intervention type* (e.g., clickers, problem-based learning, quizzing, studio/workshop). Results indicated that average examination scores did not vary significantly among subgroups of studies based on *STEM course subject matter*, *course level*, and *intervention type*. These findings suggest that the beneficial effect of active instruction does not differ significantly (a) across different STEM course subject matters, (b) whether the course is introductory (freshmen, sophomores) or upper level (juniors, seniors), or (c) based on the type of active instruction method administered. However, for *assessment type*, it was found that the average effect size was significantly lower when learning was measured with instructor-written course exams when compared with conceptual inventories. According to the authors, this finding could be explained by the different sets of cognitive skills tested by these two types of assessment. They hypothesized that it could be due to the fact that concept inventories generally test higher-level cognitive skills, and the beneficial impact of active learning is more significant for skills of that nature. For *class size*, the average effect size was found to be significantly higher for small groups, that is, for courses with 50 or fewer students. The authors did not advance any explanation with respect to this finding.

With the present study, we wish to determine whether these characteristics act as significant moderators for the learning achieved in the fields of humanities or social sciences. Thus, the following null hypotheses were tested with regard to study characteristics:

$H_{0(2)}$ : *Course subject matter* will not significantly moderate the relative difference in assessment scores between students under active learning versus students under traditional lecturing.

$H_{0(3)}$ : *Assessment type* will not significantly moderate the relative difference in assessment scores between students under active learning versus students under traditional lecturing.

$H_{0(4)}$ : *Class size* will not significantly moderate the relative difference in assessment scores between students under active learning versus students under traditional lecturing.

$H_{0(5)}$ : *Course level* will not significantly moderate the relative difference in assessment scores between students under active learning versus students under traditional lecturing.

$H_{0(6)}$ : *Intervention type* will not significantly moderate the relative difference in assessment scores between students under active learning versus students under traditional lecturing.

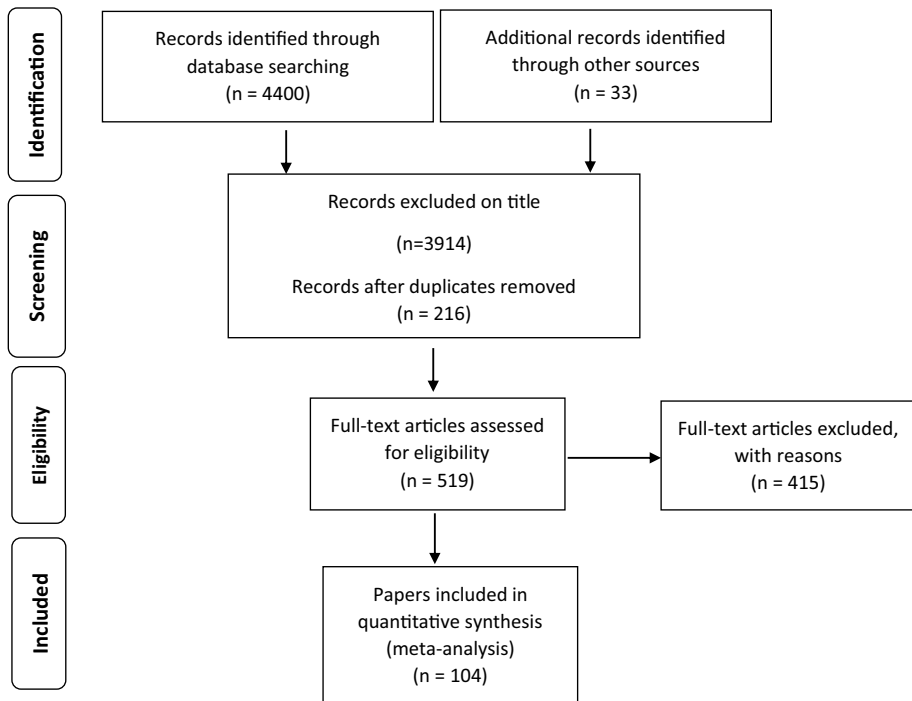
## Method

### Data collection

Three strategies were used to collect data: an electronic database search for relevant papers, a manual search through the references listed in past literature reviews and meta-analyses, and contacting individual researchers for supplementary papers.

A computer-based literature search of three databases (ERIC, Google Scholar, PsycNet) was first conducted to locate relevant studies. The database search was conducted between February 21, 2019, and October 17, 2019. Publications from January 1, 2000, through October 1, 2019, were included. The algorithm used for the search drew significant inspiration from Freeman et al. (2014) and consisted of a set of keywords related to *active learning instructional strategies* (e.g., “project-based learning,” “classroom response system,” “flipped classroom”), *post-secondary education* (e.g., “university,” “undergraduate,” “college”), *disciplines in the field of humanities* (e.g., “sociology,” “philosophy,” “politics”), *study design* (e.g., “experimental,” “quasi-experimental,” “control group”), and *learning achievement outcomes* (e.g., “learning,” “learning achievement,” “knowledge gain”). The algorithm was applied to all fields of the databases (e.g., title, abstract, descriptors, full text). To avoid an unmanageable number of results, a restriction was applied on the abstract: it could not contain any term related to a natural science discipline (e.g., “physics,” “chemistry,” “biology”). This search yielded about 4400 papers that were initially screened by reading the titles. This initial screening strategy, using a titles-first approach, was shown to be more efficient for a large corpus of papers than screening titles and abstracts together (Mateen et al., 2013). Title screening reduced the list to 486 potentially relevant papers.

In addition, a manual search was performed through the references listed in past literature reviews and meta-analyses, collected through the database search, that focused on either active learning instructional strategies as a whole or specific strategies, such as flipped classrooms, classroom response systems, or project-based learning. This manual search identified 31 additional potentially relevant papers. Lastly, individual researchers with some expertise in the field of active learning were asked to provide leads on possible supplementary papers. Their suggestions resulted in two additional potentially relevant papers, for a grand total of 519 potentially relevant papers. The following PRISMA flow diagram shows the path for evaluating and choosing the included studies (Fig. 1). PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. It is the standard for meta-analyses (Moher et al., 2009).



**Fig. 1** Flow diagram of the selection of studies

## Data evaluation

The 519 papers were evaluated for admission into the present meta-analysis through a careful reading of the abstract and by skimming the full text. Papers had to meet five criteria to be admitted:

- (1) Describe an *experimental or quasi-experimental study* that compares a group subjected to an active instructional method (e.g., project-based learning, classroom response system, flipped classroom) and a group subjected to a passive instructional method (e.g., listening to a lecture, watching a video, reading a textbook). Several papers were rejected because they were not an experiment or quasi-experiment or lacked a control group subjected to passive learning;
- (2) Focus on subject matter in the field of humanities and social sciences (e.g., sociology, philosophy, education). Several papers were rejected because they focused on subject matter in other fields, such as natural science, health science, mathematics, or engineering;
- (3) Focus on participants at a *post-secondary educational level* (e.g., undergraduates, graduates). Papers that focused on participants at primary or secondary educational levels were rejected.
- (4) Measure *learning achievement* (e.g., gain in knowledge or ability) with identical or equivalent assessments for both active and passive groups as one of the outcomes of

the study. Papers that focused only on affective or attitudinal outcomes (e.g., interest, self-efficacy, engagement) were rejected.

- (5) Provide *sufficient quantitative data* at the group level, such as means ( $M$ ), standard deviations ( $SD$ ), and sample sizes ( $N$ ), to allow for the computation of standardized effect sizes ( $d$ ,  $g$ ). Papers that examined learning achievement but provided only qualitative data were rejected. The authors of papers that provided insufficient quantitative data were contacted by email and asked to provide the missing data, if possible. When the author could not provide the data or did not reply, the paper was rejected.

Evaluation of the 519 papers was conducted by the second author of this study, an experienced coder. It was determined that 415 papers did not meet one or more of the five inclusion criteria and should be excluded from the meta-analysis. Of these 415 papers, 42 were assigned to a second coder. In 40 of the 42 cases (95% agreement), the second coder independently determined that these papers should be excluded. In the two cases of conflict, a discussion between the two coders resulted in a decision to exclude the papers from the meta-analysis. It was therefore concluded that, even with a second, independent coding, a significant number of the 415 excluded papers still would not be included in the study.

## Data synthesis

### Coding

The relevant outcome (i.e., learning achieved) and salient characteristics (i.e., hypothesized moderators) of each study were coded in a data sheet. The data that was coded and the coding system used were significantly inspired by Freeman et al. (2014). Coding of the 104 papers was conducted by the second author of this study. The second coder was randomly assigned 26 of the 104 papers and asked to independently code them. The two coders then met to discuss their respective coding of those 26 papers. Inter-rater agreement, computed as Cohen's kappa ( $\kappa$ ), was almost perfect (Cohen, 1960; McHugh, 2012) for all but three of the coded variables, with coefficients ranging between 0.86 and 1. All coding discrepancies were discussed until consensus was reached.

### Describing outcomes using a common scale (Hedges' $g$ )

The outcome of interest to the present meta-analysis (i.e., learning achieved) was computed using Hedges and Olkin's (1985) approach (see also Borenstein, 2009). The standardized mean difference effect size calculated for each study was  $d$ , being the difference between the learning achieved by the active group versus the passive group divided by the pooled standard deviation. A positive effect size therefore represents better learning achieved by the active group.

## Data analysis

### Grand mean effect size and subgroup analyses

Because it cannot be assumed that the true effect size is the same in all studies, a random-effects model was used to calculate the grand mean effect size using the

Comprehensive Meta-Analysis (CMA) 2.0 software. In this model, it is assumed that the effect sizes from each individual study are normally distributed around an overall mean. To obtain the most precise estimate of the grand mean effect size, the commonly recommended method of assigning a weight to each study and then computing a weighted mean was used (Borenstein, 2009). The weight assigned to each study was computed as the inverse of that study's variance. The term  $Q$  represents a standardized measure of weighted square deviations estimated.  $Q$  follows a central chi-squared distribution with degrees of freedom equal to  $k - 1$ . To test the null hypothesis, a  $p$ -value for the observed value of  $Q$  was thus computed with an alpha level set at 0.05. Because of the presence of significant true heterogeneity among the effect size estimates (see the “[Results](#) and [Discussion](#)” section), subgroup analyses were conducted. Then, the same procedure described for the *grand mean effect size* was applied to compute a summary effect size ( $g_w$ ) for each subgroup, a standard error on that effect size, ( $SE_{g_w}$ ), lower and upper confidence limits ( $LL_{g_w}$ ,  $UL_{g_w}$ ), and a  $p$ -value, to test the null hypothesis that the summary effect size in question was zero.

## Publication bias analyses

Studies included in a meta-analysis may overestimate the true grand mean effect size and thus bias its findings. This is because studies with significant effects or with effect sizes of greater magnitude are more likely to be published and are easier for the meta-analyst to find. This leads to a bias in published literature that can be carried over to a meta-analysis drawing on that literature. This well-documented problem for meta-analyses, referred to as the *file-drawer* problem or *publication bias*, involves drawing conclusions that are based on a biased sample of the target population of studies (Borenstein, 2009). Because publication bias is a serious threat to the validity of a meta-analysis, it must be rigorously addressed (Borenstein, 2009; 2019; Freeman et al., 2014). To do so, three analyses were performed to assess the presence of an upward bias in the grand mean effect size calculated in the present meta-analysis and to estimate how much impact this bias had: (a) assessment of funnel plots under the trim and fill method, (b) Egger's regression test (Egger et al., 1997), (c) Begg and Mazumdar's Kendall tau correlation test. The Kendall tau correlation test uses the correlation between the ranks of effect sizes and the ranks of their corresponding standard errors. A significant correlation implies publication bias. The present meta-analysis set this predetermined value to a standardized mean difference of 0.20, which is generally considered in the education literature as the smallest pedagogically significant effect size (e.g., Freeman et al., 2014; Higgins et al., 2008; Raudenbush, 2009).

## Results

### Descriptive data

The descriptive data, comprising frequency, codes, and interpretation for the salient characteristics, is available in Table 1. The 104 papers that contributed data to this meta-analysis included 15,896 students.



**Table 1** Descriptive data for salient characteristics

Characteristics	Codes and interpretation	Frequency
Course level	Intro (freshmen or sophomores)	60
	Upper (juniors or seniors)	21
	Undefined—i.e., no data on course level	30
Course subject matter	Sociology	9
	Philosophy	1
	Psychology	24
	Political science	6
	History	3
	Economics/business/finance/marketing	16
	Management	7
	Law	3
	Education	17
	Language, first language or second language	18
	Library instruction	4
	Combination (or not distinguishable)	3
	Treatment	Problem-based learning
Project-based learning		2
Case studies		2
Clicker (classroom or personal response system)		30
Flipped or Inverted classroom		36
Peer-based, team-based or collaborative learning		5
Role-play		3
Computer simulation or serious game		15
Writing assignments		2
Quizzing		2
Interactive demo		1
Experiential learning		5
Combination (two or more of the above)		6
Class size	Very small: between 1 and 20 participants	21
	Small: between 21 and 50 participants	53
	Medium: between 51 and 110 participants	20
	Large: more than 110 participants	17
Assessment type	Concept inventory	6
	Exam	95
	Course grade	8
	Other	31
Design	Presence of both a pre- and a post-test	28
	Presence of a post-test only	83
Significance	Significant increase in student learning	82
	No significant increase in student learning	29

## Main effect analyses

The weighted grand mean effect size ( $g_w$ ) representing learning achieved on identical or equivalent assessments by groups subjected to active treatment, compared with groups subjected to traditional lecturing, was a weighted standardized mean difference of 0.489 ( $Z=6.521$ ,  $p<0.001$ ,  $k=111$ ,  $N=15,896$ ) with a 95% confidence interval bounded by 0.414 and 0.564. Thus, learning achieved under active instruction increased, on average, by just slightly under half a standard deviation compared with learning achieved under traditional lecturing. As such, the primary null hypothesis of the present work must be rejected (i.e.,  $H_{0(1)}$ : Assessment scores for students under active learning will not differ significantly from assessment scores for students under traditional lecturing.)

## Subgroup analyses

The overall homogeneity analysis determined that the effect sizes were not consistent across all 111 studies, as suggested by a  $Q_T$  statistic that exceeded the critical value ( $Q_T=309.776$ ;  $df=110$ ;  $p < 0.001$ ). Moreover, computation of the  $\tau^2$  statistic yielded a value of 64.5%, meaning that there is a high percentage of real heterogeneity, or true variance, across the observed effect size estimates of individual studies. These findings led to the rejection of the null hypothesis of homogeneity across studies, indicating that there is more variability in effect sizes across studies than expected by chance alone and that it is appropriate to proceed with subgroup, or moderator, analyses.

Tables 2, 3, and 4 present the statistics computed for each of the subgroups related to each potential moderator. Findings related to moderators will follow the same order as the

**Table 2** Subgroup results by course subject matter

Variable	$k$	$N$	$\bar{g}_w$	$SE_{g_w}$	95% confidence interval limits		$Q_W$	Homogeneity of between-group effect sizes	
					Lower	Upper		$Q_B$	$p$
Course subject matter									
Sociology <sup>a</sup>	9	1467	0.640	0.178	0.291	0.989	13.776	63.478	< .001
Philosophy	1	225		0.529	-0.757	1.317	0		
Psychology <sup>a</sup>	24	5979	0.600	0.135	0.335	0.865	18.517		
Political science	6	629	0.458	0.203	0.060	0.856	27.269		
History	3	391	-0.085	0.310	-0.216	0.997	6.540		
Economics <sup>a</sup>	16	2080	0.479	0.148	0.189	0.769	12.572		
Management	7	1695	0.455	0.193	0.077	0.833	6.424		
Law	3	426	0.518	0.307	-0.084	1.120	1.123		
Education <sup>a</sup>	17	997	0.485	0.149	0.193	0.777	49.852		
Language <sup>a</sup>	18	1242	0.490	0.146	0.204	0.776	68.256		
Library instruction	4	378	0.275	0.246	-0.207	0.757	41.969		
Combination	3	387	0.175	0.292	-0.397	0.747	2.229		

<sup>a</sup>Significant pairwise difference with history, library instruction, and combination

**Table 3** Subgroup results by assessment type, group size, and course level

Variable	<i>k</i>	<i>N</i>	$\bar{g}_w$	$SE_{g_w}$	95% Confidence interval limits		$Q_w$	Homogeneity of between-group effect sizes	
					Lower	Upper		$Q_B$	<i>p</i>
Assess type									
C. invent <sup>a</sup>	6	966	0.591	0.220	0.160	1.022	1.028	17.211	< .001
Exam	95	12,671	0.485	0.108	0.302	0.712	278.421		
Grade	8	1867	0.436	0.206	0.032	0.840	1.547		
Other	31	392	0.271	0.168	0.004	0.756	19.523		
Group size									
Very small <sup>b</sup>	21	657	0.707	0.139	0.435	0.979	33.468	13.572	0.004
Small	53	4085	0.569	0.105	0.363	0.775	159.923		
Medium	20	2918	0.511	0.139	0.239	0.783	97.381		
Large	17	8236	0.372	0.149	0.080	0.664	5.431		
Course level									
Intro	60	9252	0.425	0.101	0.227	0.623	195.434	9.066	.003
Upper	21	3120	0.803	0.135	0.538	1.068	25.733		

<sup>a</sup>Significant pairwise difference with large group size

<sup>b</sup>Significant pairwise difference with other assessment types

**Table 4** Subgroup results by intervention type

Variable	<i>k</i>	<i>N</i>	$\bar{g}_w$	$SE_{g_w}$	95% confidence interval limits		$Q_w$	Homogeneity of between-group effect sizes	
					Lower	Upper		$Q_B$	<i>p</i>
Intervention									
Problem	6	430	0.634	0.218	0.207	1.061	23.255	12.96	0.29
Project	2	180	0.495	0.331	-0.154	1.144	0.296		
Case	2	377	0.563	0.379	-0.180	1.306	0.210		
Clickers	30	5785	0.408	0.127	0.159	0.657	75.991		
Flipped	36	4411	0.463	0.116	0.236	0.690	155.55		
Peer	5	393	0.515	0.222	0.080	0.950	7.084		
Role-play	2	132	0.093	0.352	-0.597	0.783	0.924		
Computer	15	1787	0.569	0.152	0.271	0.867	14.061		
Writing	2	1037	1.136	0.360	0.430	1.842	0.883		
Quizzing	2	43	1.201	0.361	0.493	1.909	0.011		
Experiential	5	636	0.655	0.221	0.222	1.088	4.914		
Combination	4	685	0.198	0.224	-0.241	0.637	13.722		

null hypotheses formulated above ( $H_{0(2)}$  to  $H_{0(6)}$ ). We remind that  $Q_W$  is the within-subgroups variance and  $Q_B$  is the between-subgroups variance.

Table 2 shows that active instruction produces significantly higher assessment scores than traditional lecturing in seven course subject matters (Sociology, Psychology, Political Science, Economics, Management, Education, Language), as the lower limit of the confidence interval on the grand mean effect size is greater than 0. Because a significant amount of heterogeneity was found across the grand mean effect sizes related to each course subject matter ( $Q_B=63.478$ ,  $p < 0.001$ ), the null hypothesis  $H_{0(2)}$  was rejected. A pairwise comparison of the mean effect size estimates related to each subject matter yielded several significant pairwise differences.

Table 3 first shows that active instruction produces significantly higher assessment scores than traditional lecturing across the four types of assessments considered (i.e., concept inventories, instructor or researcher-written exams, course grades and others [e.g., quizzes, assignments]). Moreover, because a significant amount of heterogeneity was found across the grand mean effect sizes related to each type of assessment ( $Q_B = 17.999$ ,  $p < 0.001$ ), the null hypothesis  $H_{0(3)}$  was rejected. This finding suggests that active instruction benefits students regardless of the type of assessment used to measure learning achievement. A pairwise comparison of the mean effect size estimates related to each type of assessment yielded one significant pairwise difference, between concept inventories, the assessment type with the highest mean effect size estimate, and the assessment types qualified as other (e.g., quizzes, assignments), which had the lowest mean effect size estimate.

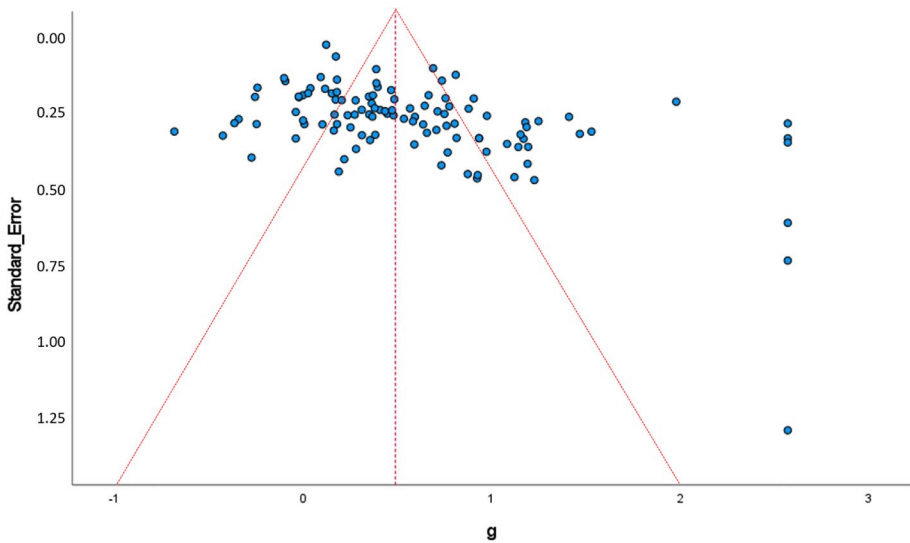
As for group size, Table 3 shows that active instruction produces significantly higher assessment scores than traditional lecturing through all group sizes. This finding suggests that active instruction benefits students regardless of group size. Because a significant amount of heterogeneity was found across the grand mean effect sizes related to each group size ( $Q_B = 13.572$ ,  $p = 0.004$ ), the null hypothesis  $H_{0(4)}$  was rejected.

Finally, Table 3 shows that active instruction produces significantly higher assessment scores than traditional lecturing for both introductory- (i.e., freshman, sophomore) and upper-level (i.e., junior, senior) courses. This finding suggests that active instruction benefits students regardless of course level. Because a significant amount of heterogeneity was found between the grand mean effect sizes related to introductory- and upper-level courses ( $Q_B = 9.066$ ,  $p = 0.003$ ), the null hypothesis  $H_{0(5)}$  was rejected. The significantly higher mean effect size associated with upper-level courses does not concur with Freeman et al.'s meta-analysis, which found a non-significant difference between introductory- and upper-level courses.

Table 4 shows that eight of the twelve types of active treatments produce significantly higher assessment scores than traditional lecturing (i.e., problem-based, clickers, flipped, peer-based, computer-based, writing, quizzing, experiential). The other four types of active treatment do not produce higher assessment scores relative to traditional lecturing (i.e., project-based, case study, role-play, combination). Because the amount of heterogeneity between the grand mean effect sizes across active treatment categories was not found to be statistically significant ( $Q_B = 12.959$ ,  $p = 0.296$ ), the null hypothesis  $H_{0(6)}$  was accepted.

## Publication bias findings

Figure 2 shows the funnel plot, with standard error on effect size estimates ( $SE_g$ s) as a function of effect size estimates ( $g$ s). Visual inspection of the funnel plot indicates some asymmetry in the data, due to six or seven data points with unusually large effect size



**Fig. 2** Funnel plot illustrating the standard error on effect size estimates (*SEgs*) as a function of effect size estimates (*gs*)

estimates (*gs*) and/or standard errors (*SEgs*). Moreover, both Egger’s regression test and Kendall tau rank correlation test indicate a statistically significant association between effect size estimates and their corresponding standard errors. Kendall tau coefficient with continuity correction yielded a value of 0.330 with a one-tailed *p*-value of  $< 0.001$ . Egger’s regression intercept yielded a value of 0.180 with a one-tailed *p*-value of 0.007. Both tests thus confirm there is asymmetry in the funnel plot.

The “adjusted” estimate of the grand mean effect size calculated under the trim and fill method was slightly lower than the one obtained with the original analysis ( $g_{\text{wadj}} = 0.476$ ) with a 95% confidence interval bounded by 0.401 and 0.551. Thus, the adjusted estimate is very close to the original estimate, with a 0.476 estimate having the same substantive implications as the original estimate of 0.489. The trim and fill analysis indicates that the degree of asymmetry observed in the present meta-analysis has virtually no impact on the estimate of the grand mean effect size. This finding indicates that bringing the grand mean effect size down to a value that would be pedagogically insignificant would require an unreasonably large number of studies with a null effect that went undetected. As such, there is no indication that publication bias has affected the grand mean effect size estimate reported in the present meta-analysis.

## Discussion

The present work, inspired by Freeman et al. (2014), sought to examine the effect of active instruction on learning achievement for college programs in the field of humanities and social sciences. To that end, we meta-analyzed 104 studies that used assessment scores to compare the learning achieved by groups of college students in humanities and social science disciplines under active learning methods versus traditional lecturing. The weighted grand mean effect size estimated from these studies indicates that student performance on

assessment scores is on average higher under active instruction, and is similar to the grand mean effect size in Freeman et al.'s study. This similarity suggests that the magnitude of the beneficial effect on learning achieved in higher education through active instruction, compared with traditional lecturing, is similar in both STEM and the humanities and social sciences. This conclusion is further supported by several earlier meta-analyses on the impact of active alternatives versus traditional lecturing on the learning achieved by college students, which reported similar magnitudes of grand mean effect size estimates in favor of active instruction (Dochy et al., 2003; Ruiz-Primo et al., 2011; Swanson et al., 2017; Shi et al., 2019).

With regard to practical implications, the grand mean effect size reported in the present meta-analysis suggests that active instruction should be either maintained (if already implemented) or strongly considered by policymakers and teachers as at least a partial replacement for traditional lecture-based instruction in college programs in the fields of humanities and social sciences. Institutions and policymakers should encourage their instructors to adopt active teaching practices, drawing on and emphasizing the scientific value of the results obtained through studies such as this one.

Subgroup analyses found that four variables had a significant moderator effect on the grand mean effect size estimate: *course subject matter*, *assessment type*, *class or group size*, and *course level*. First, learning achieved with active instruction was found to be most beneficial in some humanities and social sciences course subject matters, specifically Sociology, Psychology, Language, Education, and Economics. This finding suggests that faculties offering programs in these disciplines should seriously consider replacing traditional lecturing with active instruction. It cannot be concluded that active instruction benefits learning for the five other course subject matters (Philosophy, History, Law, Library instruction, Combination). However, this could be due to lack of studies for these five course subject matters, so significance is harder to establish, given the low levels of homogeneity of between-group effect sizes. Therefore, we must be careful not to conclude hastily that active learning instruction methods would not be beneficial for these course subject matters, given the very low number of studies. Also, significant pairwise difference was found between Sociology, the course subject matter with the highest mean effect size estimate, and the course subject matters with the three lowest mean effect size estimates (i.e., History, Library instruction, Combination), but not Philosophy (fourth lowest), most likely due to the fact that only one study was included in Philosophy. Consequently, further research is required to develop a better understanding of the effect active instruction has on learning achievement in the aforementioned college course subject matters treated here.

Second, the findings of the present work regarding group size suggest that learning achieved with active instruction was found to be higher for all group sizes. This finding does not concur with Freeman et al. (2014) and other earlier meta-analyses (Shi et al., 2019; Swanson et al., 2017), which found that learning achieved with active instruction is greater with smaller group sizes. This discrepancy between our findings and previous studies carried out within STEM fields seems to support the limits of transferring research finding across disciplines (Adler et al., 2018; Bartha, 2013). That said, a pairwise difference did show that learning achieved with active instruction is greater with very small class or group sizes, compared with large class or group sizes for programs in the humanities and social sciences. There is a plethora of empirical data and conceptual arguments in the literature regarding mechanisms that explain why active instruction is more beneficial with smaller classes in the context of higher education. These include higher levels of student engagement, increased time spent on tasks and the opportunity for teachers to maintain higher quality personal interactions with their

students, better tailor instruction to their ability levels and interests, and better monitor their progress (Ballen et al., 2018, Ho & Kelman, 2014, Schanzenbach, 2014, Baker et al., 2016). Such confounding variables should be considered in future research.

Third, learning achieved with active instruction was found to be higher for both introductory-level and upper-level courses as compared to traditional instruction. Although this moderator has not been frequently examined by earlier meta-analyses, some have found that active instruction appears to be more beneficial for students in upper-level courses. For example, Dochy et al.'s meta-analysis (2019), which compared the effect of problem-based learning with traditional lecturing for medical students, found greater effect sizes for students in the final 2 years of their program ( $ES = 0.732$  and  $0.679$ ), as compared to students in the first 2 years ( $ES = 0.414$  and  $0.473$ ), with respect to skill outcomes. Our results seem to agree with such previous findings, where the grand mean effect sizes were higher for upper-level courses. There are several reasons that could help explain the increased benefit of active instruction for upper-level courses. These courses generally mobilize higher-level cognitive skills (i.e., problem-solving) as compared to introductory-level courses, which tend to mobilize more content-mastery skills. Previous research (e.g., Atman et al., 2005; Tsenn et al., 2013) has also shown that juniors and seniors generally possess a higher level of self-efficacy and better transversal competencies than freshmen and sophomores. Because these factors are associated with higher academic achievement, they could explain the findings observed here. Lastly, it is important to reiterate that 30 of 104 studies included in the present meta-analysis did not report information on course level. Consequently, one recommendation for future research on active learning in the field of humanities and social sciences would be to record data more assiduously regarding course level. This would allow for a more accurate estimate of this variable as a potential moderator of the learning achieved.

Subgroup analyses associated with the *type* of active treatment implemented, identified as potentially impacting the grand mean effect size, did not however yield significant results, suggesting that this variable does not act as moderator. This finding suggests that the beneficial effect of active instruction does not vary significantly with the type of active treatment implemented. This concurs with Freeman et al.'s meta-analysis, which did not find a significant difference across active treatment categories. However, it is interesting to note that some earlier meta-analyses (e.g., Ruiz-Primo et al., 2011; Schroeder et al., 2007) found that different types of active instruction yielded significantly different results on the learning achieved. Because there appears to be an important disparity in the mean effect size estimates associated with each active treatment, this finding could be due to the fact that several active treatment categories included a very low number of studies, resulting in a lack of statistical power and precluding more robust findings. For example, five categories (project, case study, role-play, quizzing, writing) included only two studies. Some pairwise differences related to these categories appeared to approach statistical significance but did not reach it, likely because of the small number of studies included (e.g., quizzing vs. clickers, writing vs. clickers, quizzing vs. project, writing vs. project). For these reasons, the non-significant finding reported here with regard to *intervention type* must be interpreted with a great deal of caution. As such, further primary studies are needed to gain a better understanding of the effect of these latter active treatments on the learning achieved in humanities and social science college programs.

## Conclusion

This meta-analysis is one of the first to exclusively consider comparing active and passive instruction practices in the fields of study in the humanities and social sciences. Results reported here provide sound scientific evidence for the overall superiority of active instruction for learning achievements in the context of higher education, as compared to passive instruction. These results are in line with previous similar studies conducted in the fields of science, technology, engineering, and mathematics (STEM). Moreover, findings suggest that this effect is beneficial with any type of active teaching method. By considering various variables that may act as moderators, we found the following four to have a significant effect: course subject matter, assessment type, group size, and course level. Specifically, small group sizes are the ones benefiting the most, with large group sizes benefiting the least. Upper-level courses also seem to benefit the most from active learning methods. Regarding the differences between some course subject matters considered in this study, we must interpret the results with caution as some are under-represented. It is the case for Philosophy, History, Law, and Library instruction. Thus, more primary studies are needed in these course subject matters to eventually help increase the homogeneity of between-group effect sizes. Nevertheless, institutions and policymakers should encourage their instructors to adopt active teaching methods in the humanities and social sciences.

**Funding** Our work was funded by Grant 252996 from the Fonds de Recherche Société et Culture Québec. We would also like to thank the Centre de Recherche Interuniversitaire sur la Formation et la Profession Enseignante, Université du Québec, for their financial support.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adler, C., Hirsch Hadorn, G., Brey, T., Wiesmann, U., & Pohl, C. (2018). Conceptualizing the transfer of knowledge across cases in transdisciplinary research. *Sustainability Science*, *13*, 179–190. <https://doi.org/10.1007/s11625-017-0444-2>
- Atman, C. J., Cardella, M., Turns, J., & Adams, R. (2005). Comparing freshman and senior engineering design processes: An in-depth follow-up study. *Design Studies*, *26*(4), 325–357. <https://doi.org/10.1016/j.destud.2004.09.005>
- Baker, B. D., Farrie, D., & Sciarra, D. G. (2016). Mind the gap: 20 years of progress and retrenchment in school funding and achievement gaps. *Educational Testing Service*. <https://doi.org/10.1002/ets2.12098>
- Ballen, C. J., Aguillon, S. M., Brunelli, R., Drake, A. G., Wassenberg, D., Weiss, S. L., ... Cotner, S. (2018). Do small classes in higher education reduce performance gaps in STEM? *BioScience*, *68*, 593–600. <https://doi.org/10.1093/biosci/biy056>
- Barkley, E. F., Major, C. H., & Cross, P. K. (2014). *Collaborative learning techniques*. John Wiley & Sons.



- Barkley, E.F. (2009). *Student engagement techniques: A handbook for college faculty*. John Wiley & Sons.
- Bartha, P. (2013). Analogy and analogical reasoning. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/reasoning-analogy/>. Accessed July 2021
- Biggs, J. (2003). *Teaching for quality learning at university*. Buckingham: Open University Press.
- Bishop, J. L. & Verleger, M. A. (2012). *The flipped classroom: A survey of the research*. [Paper presentation]. American Society of Engineering Education 120<sup>th</sup> Annual Conference & Exposition, Atlanta, Georgia. [https://www.researchgate.net/publication/285935974\\_The\\_flipped\\_classroom\\_A\\_survey\\_of\\_the\\_research](https://www.researchgate.net/publication/285935974_The_flipped_classroom_A_survey_of_the_research). Accessed July 2021
- Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis*, 2<sup>nd</sup> edition (pp. 221–235). New York, NY: Russell Sage Foundation. <https://psycnet.apa.org/record/2009-05060-012>. Accessed July 2021
- Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them*. Biostats (409 p). <https://meta-analysis-books.com/>. Accessed May 2022
- Brunner, R. D. (2010). Adaptive governance as a reform strategy. *Policy Sciences*, 43(4), 301–341. <https://doi.org/10.1007/s11077-010-9117-z>
- Burgess, A. W., McGregor, D. M., & Mellis, C. M. (2014). Applying established guidelines to team-based learning programs in medical schools: A systematic review. *Academic Medicine*, 89(4), 678–688.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. <https://doi.org/10.1177/001316446002000104>
- Condliffe, B., Quint, J., Visher, M. G., Bangser, M. R., Drohojowska, S., Saco, L., & Nelson, E. (2017). Project-based learning: A literature review. <https://files.eric.ed.gov/fulltext/ED578933.pdf>. Accessed July 2021
- Deschryver, N., & Lameul, G. (2016). Vers une opérationnalisation de la notion de posture professionnelle en pédagogie universitaire. *Revue Internationale De Pédagogie De L'enseignement Supérieur*, 32(3), 1–15.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629–634.
- Fies, C., & Marshall, J. (2006). Classroom response systems: A review of the literature. *Journal of Science Education and Technology*, 15(1), 101–109. <https://doi.org/10.1007/s10956-006-0360-1>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences (PNAS)*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Grosemans, I., Coertjens, L., & Kyndt, E. (2017). Exploring learning and fit in the transition from higher education to the labour market: A systematic review. *Educational Research and Reviews*, 21, 67–84. <https://doi.org/10.1016/j.edurev.2017.03.001>
- Hartikainen, S., Rintala, H., Pylväs, L., & Nokelainen, P. (2019). The concept of active learning and the measurement of learning outcomes: A review of research in engineering higher education. *Educational Science*, 9, 276. <https://doi.org/10.3390/educsci9040276>
- Higgins, J. P. T., Deeks, J. J., & Altman, D. G. (2008). Special topics in statistics. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic review of interventions* (pp. 481–529). Hoboken, NJ: Wiley. <https://training.cochrane.org/handbook/current>. Accessed July 2021
- Hill, J., & Jordan, L. (2021). Instructional Strategies. In J.K. McDonald & R.E. West (Eds.), *Design for learning: Principals, Processes, and Praxis*. EdTech Books. [http://edtechbooks.org/id/instructional\\_strate](http://edtechbooks.org/id/instructional_strate). Accessed July 2021
- Ho, D. E., & Kelman, M. G. (2014). Does class size affect the gender gap? A natural experiment in law. *Journal of Legal Studies*, 43(2), 291–321. <https://doi.org/10.1086/676953>
- Hofer, S., Schumaker, R., Rubin, H., & Stern, E. (2018). Enhancing physics learning with cognitively activating instruction: A quasi-experimental classroom intervention study. *Journal of Educational Psychology*, 110(8), 1–17. <https://doi.org/10.1037/edu0000266>
- John, A., Shahzadi, D., & Khan, K. (2016). Students' preferred learning styles and academic performance. *Science International*, 28(4), 337–341.
- Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools*, 19(3), 267–277. <https://doi.org/10.1177/1365480216659733>
- Kozanitis, A. & Quévillon-Lacasse, C. (2018). Étude exploratoire de l'utilisation des TICE en soutien aux pédagogies actives en contexte d'enseignement universitaire. *Méditations et Méditatisations. Revue internationale sur le numérique en éducation et communication*. 1(1), 50–71. Online: <https://revue-meditations.teluq.ca/index.php/Distances/article/view/57/25>. Accessed July 2021
- Lima, R. M., Andersson, P. H., & Saalman, E. (2017). Active learning in engineering education: A (re)introduction. *European Journal of Engineering Education*, 42, 1–4. <https://doi.org/10.1080/03043797.2016.1254161>

- Mateen, F. J., Oh, J., Tergas, A. I., Bhayani, N. H., & Kamdar, B. B. (2013). Titles versus titles and abstracts for initial screening of articles for systematic reviews. *Clinical Epidemiology*, *5*, 89–95. <https://doi.org/10.2147/CLEP.S43118>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica*, *22*(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education*, *30*, 159–167. <https://doi.org/10.1152/advan.00053.2006>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISME statement. *PLoS Med*. <https://doi.org/10.1371/journal.pmed.1000097>
- Newman, M. (2003). *Pilot systematic review and meta-analysis on the effectiveness of problem based learning*. [https://www.researchgate.net/publication/242422073\\_A\\_Pilot\\_Systematic\\_Review\\_and\\_Meta-Analysis\\_on\\_the\\_Effectiveness\\_of\\_Problem-Based\\_Learning](https://www.researchgate.net/publication/242422073_A_Pilot_Systematic_Review_and_Meta-Analysis_on_the_Effectiveness_of_Problem-Based_Learning). Accessed July 2021
- Patton, C. M. (2015). Employing active learning strategies to become the facilitator, not the authoritarian: A literature review. *Journal of Instructional Research*, *4*, 134–141. <https://eric.ed.gov/?id=EJ1127732>. Accessed July 2021
- Pratt, D. (1992). Conceptions of Teaching. *Adult Education Quarterly*, *42*(4), 203–220.
- Prince, M. J. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*, 223–231. [https://www.engr.ncsu.edu/wp-content/uploads/drive/1smSpn4AiHSh8z7a0MHDBwhb\\_JhcoLQml/2004-Prince\\_AL.pdf](https://www.engr.ncsu.edu/wp-content/uploads/drive/1smSpn4AiHSh8z7a0MHDBwhb_JhcoLQml/2004-Prince_AL.pdf). Accessed July 2021
- Rahman, A. A., Mohamed, H., Aris, B., & Zaid, N. M. (2014). The influences of flipped classroom: A meta analysis. [Paper presentation], *Institute of Electrical and Electronics Engineers 6th International Conference on Engineering Education*, Kuala Lumpur, Malaysia. <https://doi.org/10.1109/ICEED.2014.7194682>
- Rana, N. P., Dwivedi, Y. K., & Al-Khowaiter, W. A. A. (2016). A review of literature on the use of clickers in the business and management discipline. *The International Journal of Management Education*, *14*(2), 74–91. <https://doi.org/10.1016/j.ijme.2016.02.002>
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*, 2<sup>nd</sup> edition, pp. 295–315. New York, NY: Russell Sage Foundation. <https://psycnet.apa.org/record/2009-05060-012>. Accessed July 2021
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, *331*(6022), 1269–1270. <https://doi.org/10.1126/science.1198976>
- Schanzenbach, D. W. (2014). *Does Class Size Matter?* Boulder, CO: National Education Policy Center. <http://nepc.colorado.edu/publication/does-class-size-matter>. Accessed July 2021
- Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T. Y., & Lee, Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, *44*, 1436–1460. <https://doi.org/10.1002/tea.20212>
- Shi, Y., MacLeod, J., & Hao Yang, H. (2019). College students' cognitive learning outcomes in flipped classroom instruction: A meta-analysis of the empirical literature. *Journal of Computers in Education*, *7*, 79–103. <https://doi.org/10.1007/s40692-019-00142-8>
- Simpson, V., & Oliver, M. (2006). *Using electronic voting systems in lectures*. <https://pdfs.semanticscholar.org/ffc4/fe13e896b198f72fa3a43a04e383a5d78d51.pdf>. Accessed July 2021
- Swanson, E., McCulley, L. V., Osman, D. J., Scammacca Lewis, N., & Solis, M. (2017). The effect of team-based learning on content knowledge: A meta-analysis. *Active Learning in Higher Education*, *20*(1), 1–12. <https://doi.org/10.1177/1469787417731201>
- Thomas, J.W. (2000). *A review of research on project-based learning*. California: The Autodesk Foundation. [http://www.bobpearlman.org/BestPractices/PBL\\_Research.pdf](http://www.bobpearlman.org/BestPractices/PBL_Research.pdf). Accessed July 2021
- Tormey, R. (2021). Rethinking student-teacher relationships in higher education: A multidimensional approach. *Higher Education*, *82*, 993–1011.
- Tsenn, J., McAdams, D. A., & Linsey, J. S. (2013). A Comparison of Design Self-Efficacy of Mechanical Engineering Freshmen, Sophomores, and Seniors. [Paper presentation]. American Society for Engineering Education 120<sup>th</sup> Annual Meeting, Atlanta, Georgia, United States. <https://peer.asee.org/a-comparison-of-design-self-efficacy-of-mechanical-engineering-freshmen-sophomores-and-seniors.pdf>.
- Uzunboylu, H., & Karagozlu, D. (2015). Flipped classroom: A review of recent literature. *World Journal on Educational Technology*, *7*(2), 142–147. <https://doi.org/10.18844/wjjet.v7i2.46>