Student-Centered Learning in Undergraduate Level Science Post-Secondary Education and Academic Achievement: A Meta-Analysis

Brian Mihov

A Thesis
In the Department
of
Education

Presented in Partial Fulfillment of the Requirements
For the Degree of
Master of Arts (Educational Technology) at
Concordia University
Montreal, Quebec, Canada

October 2019

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Brian Mihov

Entitled: Student-Centered Learning in Undergraduate Level Science Post-Secondary Education and Academic Achievement: A Meta-Analysis

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Educational Technology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

______________________________ Chair

Dr. Julie A. Corrigan

______________________________ Examiner

Dr. Robert M. Bernard

______________________________ Examiner

Dr. Eugene Borokhovski

______________________________ Supervisor

Dr. Richard Schmid

Approved by ______ Dr. Sara Kennedy ______
Chair of Department or Graduate Program Director

October 2 2019 ______ André G. Roy ______
Dean of Faculty of Arts and Science

Abstract

Student-Centered Learning in Undergraduate Level Science Post-Secondary Education and Academic Achievement: A Meta-Analysis

Brian Mihov

This meta-analysis assesses the overall impact on undergraduate level science post-secondary student achievement outcomes of instructional environments that are more student-centered versus less student-centered (more teacher-centered). It also considers in which of four instructional events (dimensions) – *Pacing, Teacher's Role, Flexibility and Adaptation* – the application of more student-centered pedagogy is more optimal for increasing student achievement outcomes, as well as considers the strength of student-centered pedagogy in each of these four instructional dimensions. Additionally, this meta-analysis considers the impact of a set of instructional and demographic moderator variables – technology use, subject matter, and treatment group class size – on student achievement. Out of an initial pool of 9759 abstracts, 96 full-text sources were chosen for analysis, yielding 141 independent effect sizes. The random effects model weighted average effect size was $\overline{g} = 0.34$, $k = 141$, $SE = 0.04$, $z = 8.58$, $p < .001$, suggesting that on average *more* student-centered classroom studies produce better results on achievement outcomes than do *less* student-centered classroom studies. However, the non-significant meta-regression result ($p = 0.40$) compromises the strength of this conclusion. Of the four instructional dimensions, based on simple meta-regression, only *Flexibility* produced a significant (negative) relationship ($\beta = -0.09$, $p \leq .05$). Mixed moderator variable analysis yielded the subject matter of chemistry ($\beta = 0.23$, $p = 0.03$)

as the best predictor of effect size; studies in which both groups used technology had a significantly lower average effect size (β = -0.18, p = 0.04) than the reference group of studies in which both groups did not use technology; and in studies in which the treatment group used technology and the control group did not, the result was not significantly different from studies in which both groups did not use technology (β = 0.07, p = 0.53). Recommendations include attending to more nuanced moderator variables when introducing student-centered strategies.

Table of Contents

List of Tables

List of Figures

Background

In order to provide students with optimal opportunities to learn and to apply knowledge, the appropriate educational environments need to be present. Two particular environments, teacher-centered (T-C) and student-centered (S-C) learning, have enjoyed extensive research and application across K-12 and post-secondary education. Additionally, a decades-spanning dichotomy has existed between these two approaches to learning, with a large body of literature characterized by an "either/or" stance.

T-C learning, alternatively referred to as the more traditional approach to instruction, sees the instructor assuming the main responsibilities regarding course planning and objectives, as well as applying more direct instructional methods. Observational studies of classroom instruction, such as the research conducted by Rosenshine and Stevens (1986), explored correlations between teacher behaviour and student achievement outcomes, and identified instructional behaviours and patterns that ran parallel to those described via terms such as "direct instruction" (Rosenshine, 1976), "explicit teaching" (Stanovich, 1980), and "systematic teaching" (Morrison, 1926). Specifically, Rosenshine and Stevens (1986) described and grouped their results into six sequential teaching functions reflecting a direct instruction approach: 1) daily review; 2) presenting new material; 3) guiding student practice; 4) providing feedback and corrections; 5) conducting independent practice; and 6) weekly and monthly reviews. Throughout the T-C instructional process, students' learning experience flows through the instructional conditions and parameters implemented by the teacher.

S-C learning, on the other hand, sees the teacher taking on more of a facilitator role. More indirect instructional methods are implemented, as individual students, small groups of students or students within classrooms as a whole, experience a more individualized learning environment. Jonassen (1991) proposed S-C instructional design principles that included: 1) creating real-world environments that employ the context in which learning is relevant; 2) focusing on realistic approaches to solving real-world problems with the instructor acting as a coach and analyzer of the strategies used to solve these problems; 3) stressing conceptual interrelatedness, providing multiple representations or perspectives on the content; 4) defining instructional goals and objectives as negotiated and not imposed; 5) designing evaluations that serve as a self-analysis tool; 6) providing tools and environments that help students interpret multiple perspectives upon the world; and 7) favouring learning that is internally controlled and mediated by the student.

While T-C learning has been depicted as representing the predominant approach, S-C learning has been present and exerting influence in research and in classrooms for some time. As an educational paradigm, S-C learning has roots in the writings and ideas of influential educators such as John Dewey (1938), Lev Vygotsky (1962, 1978) and others, as well as roots in constructivist learning theories. A main tenet of constructivist theory, as it pertains to learning, is that knowledge is situated in the activity of the learner, and that knowledge is a product not only of that activity, but also of the context and culture in which it occurs (Brown, Collins, & Duguid, 1989). Notable literature on the S-C approach to learning includes the twenty-nine school, eight-year study on the effectiveness of progressive education (Aiken, 1942), as well as the Plowden Report (1967) in Britain advocating a child-centered approach to primary education.

In terms of notable literature on T-C learning, the U.S. President Lyndon Johnson administration's 1960's *Project Follow Through* was a large-scale research project exploring the effectiveness of educational strategies spanning from direct instruction to open education. The main findings centered on direct instruction outperforming other models on achievement and affect measures by as much as 1.5*sd*. The findings, which, along with the conducting of the study, were seen as controversial, nevertheless were instrumental in positioning T-C learning as the predominant learning approach in educational circles. While *Project Follow Through* remains an influential research project advocating for direct instruction, other work, such as Brophy and Good (1986) explored pretest-posttest designs of classroom instruction, and the classes in which students attained the highest achievement gains were found to be led by teachers exhibiting more of a direct instructional approach. It should be noted that the assessments were typically in reading and mathematics. Other research has looked into drill and practice programs, including a meta-analysis on computer-based instruction (CBI) by Kulik (1994) exploring drill and practice CBI programs on arithmetic, spelling, and vocabulary in beginning reading. The questions all had one unambiguous correct answer, and effect sizes for correct answers were shown to be 0.50 and higher (Kulik, 1994).

While the above literature examples pertain more to an "either/or" approach, there has been literature opting for a more non-mutually exclusive approach by attempting to identify combinations of T-C and S-C learning in educational settings. A prominent example is Gersten et al. (2008), who conducted a meta-analysis of T-C and S-C mathematics teaching practices. Their concluding remarks reflected on the fact that no studies contained examples of students teaching themselves or others without any teacher guidance, nor were there any examples of teachers not taking into account, or not paying attention to, students' understanding and responses to the content being taught (Gersten et al., 2008). Similar findings were also reported in the 2008 National Mathematics Advisory Panel Final Report, and which also noted that the approach of teachers was more representative of a blended approach of T-C and S-C, rather than a distinct "either/or" approach (National Mathematics Advisory Panel, 2008). Gersten et al.

(2008) also pointed to the role of operationalization in the inability to ascertain which instructional approach is superior by stating that, "the fact that these [T-C and S-C] terms, in practice, are neither clearly nor uniformly defined, nor are they true opposites, complicates the challenge of providing a review and synthesis of the literature…" (p. 12).

The concluding remarks by Gersten et al. (2008) additionally point to the variety of instructional events taking place in a given instructional setting, as well as to how these different instructional events require different learning approaches as a function of the instructional environment.

In this project, instructional events (also referred throughout the project as 'instructional dimensions') are different aspects within an instructional environment that are combined to characterize the teaching and learning processes occurring.

In an attempt to capture the combinations of T-C and S-C learning methods occurring in instructional environments, Bernard et al. (2013) developed a literature-based set of 11 instructional events (dimensions) that can be found in a given instructional environment, as well as a five-point scale ranging from 1 (mostly T-C learning) to 5 (mostly S-C learning) in order to code each instructional dimension separately on amount of T-C/S-C learning. These 11 dimensions can serve as moderator variables for student learning achievement outcomes, and their scores can be added in order to determine the overall relationship between T-C and S-C learning for a given instructional environment.

Bernard et al. (2013) identified the following 11 instructional dimensions:

Instructional Dimensions	Description of Dimensions
Course Design	Degree to which teachers/students participate in course design.

Learning Objectives	Degree to which teachers/students set learning objectives.
Learning Materials	Degree to which teachers/students select/prepare study materials.
Adaptation of Materials	Depicts the extent to which materials and learning activities are generic and unmodified or individualized to account for differences in students' interests and abilities.
Pacing of Instruction	Degree to which students are involved in determining the pace of instruction/learning activities.
Anchored Instruction	Degree to which the instruction/exercises are authentic or anchored in realistic scenarios.
Type of Problems	Specifies cognitive processes tapped for successfully solving different problems – roughly corresponds to Bloom's Taxonomy – from well-structured algorithmic tasks to ill-structured creative problem-solving.
Conceptual Level	Describes cognitive/meta-cognitive level of the objectives being achieved (e.g., memorization, analyses, understanding, explanation, self-regulation).
Teacher's Role	Describes the teacher's role in the classroom (from lecturer and authority figure to facilitator and partner).
Peer Collaboration	Identifies the extent to which students work collaboratively/cooperatively in groups or

	teams.
Peer assessment	Degree to which students participate in feedback provision and assessment of each other's learning.

With not only a large number of, but also interactions among, different moderator variables for the amount and type of T-C and S-C learning in instructional settings and how they subsequently influence student learning achievement outcomes, an overarching approach that views and deconstructs teaching and learning according to the events associated with instructional conditions is required (Bernard et al., 2019). Such an approach veers away from an "either/or" view of T-C and S-C learning, and towards a "greater-to-lesser" S-C learning scale along a continuum of instructional practices. By examining instructional events in isolation, a greater understanding of the interactions among those events is possible, and a subsequent greater understanding of what combinations of T-C and S-C learning will lead to optimal learning environments also becomes more probable.

Bernard, Borokhovski, Schmid, Waddington, & Pickup (2019) applied a process of examining K-12 instruction practices and processes as a combination of T-C and S-C methods via a meta-analysis that sought to: 1) summarize research on the effectiveness (learning achievement outcomes) of more S-C pedagogical approaches; and 2) explore what combinations of T-C and S-C qualities of instructional interventions influence learning outcomes the most and what moderator variables either improve or reduce these learning outcomes. The authors used a literature-based set of 4 instructional dimensions (*Pacing*, *Teacher's Role*, *Flexibility*, and *Adaptation*), taken from the 11 instructional dimensions developed by Bernard et al. (2013), and coded them on a five-point scale ranging from predominantly T-C to predominantly S-C. They applied these dimensions and coding to experimental studies (quasi-experimental and random assignment) and extracted effect sizes for achievement outcomes (k = 365) from those studies where the treatment condition(s) were coded as more S-C. An average weighted effect size of 0.444*sd* was found, indicating that, on average, classroom studies with more S-C learning produce better results on achievement outcomes than do classroom studies with less S-C learning (Bernard et al., 2019). Additionally, meta-regression analysis of the four instructional dimensions resulted in two of the four dimensions (*Pacing* and *Teacher's Role*) being significant predictors of effect size. An interesting finding was that *Pacing* was a negative predictor of effect size, while *Teacher's Role* was positive.

The authors also explored a number of substantive and demographic moderator variables in an attempt to further learn about the conditions under which higher learning outcomes occur. Among the substantive moderator variables were treatment duration, instructor's experience, provision of professional development for teachers, and training for students, and demographic moderator variables included learners' age, educational background and ability level, and subject matter studied (Bernard et al., 2019). Mixed moderator variable analysis for demographics found only one significant difference – *Ability Profile* – with an effect size of 0.42*sd* (k = 338) for *General Population*, and 0.80*sd* (k = 26) for *Special Education*. The only other demographic variable that was close to a significant difference was *Subject Matter*, with an effect size of 0.40*sd* (k = 260) for *STEM* and 0.52*sd* (k = 93) for *Non-STEM*.

While the meta-analysis by Bernard et al. (2019) focused on a variety of instructional events and additional substantive and demographic moderator variables that influence learning achievement outcomes as a function of the amount and type of T-C and S-C learning in K-12, undergraduate level post-secondary instructional environments also include similar, and different, moderator variables, which in turn also influence the “greater-to-lesser” S-C learning continuum found across these particular instructional settings. Like in K-12 settings, these variables and their combinations, along with the overall degree of S-C learning, also have the ability to influence achievement outcomes in undergraduate level post-secondary settings.

Post-secondary education is education after high school, and for the purposes of this project only undergraduate level post-secondary education leading to a Bachelor’s degree will be explored, also by means of meta-analysis. More specifically, this meta-analysis seeks to explore undergraduate level science education and its relationship with T-C and S-C learning. While there are numerous subject matters that fall under the “science” umbrella, for the purposes of operationalization this meta-analysis will delve into five specific subject matters: chemistry, physics, biology, geology and psychology (clinical and experimental). Regarding the literature advocating for a spectrum-based approach to T-C and S-C learning, the subject matter of undergraduate level science has not been exclusively researched at a systematic review level. Aiello and Wolfle (1980) did conduct a 30-study meta-analysis on individualized instruction in science compared with a traditional lecture method, however individualized instruction was operationalized and separated into particular methods of instruction (audio-tutorial, computer-assisted, personalized system, programmed, and a combination category) rather than into particular instructional dimensions. Furthermore, as has been the general theme regarding T-C/S-C literature, the meta-analysis by Aiello and Wolfle (1980) was more centered around which method of instruction was superior, rather than framing the research questions from a “which,” “when,” and “for what purpose” perspective.

To once again reference the remarks by Gersten et al. (2008), there are a variety of instructional events taking place in a given instructional setting, and this same notion can be applied to an undergraduate level instructional setting, and more specifically to an undergraduate level science setting. Undergraduate level science post-secondary instructional settings are also made up of various instructional events, including the four dimensions (*Pacing*, *Teacher's Role*, *Flexibility*, and *Adaptation*) explored by Bernard et al. (2019). Like in K-12, the amount and type of T-C and S-C learning in each dimension is a function of the given instructional environment being explored.

A number of additional moderator variables – both instructional and demographic – fall under the undergraduate level science post-secondary education umbrella, and as a result a better understanding of how primary predictor moderator variables (*Pacing*, *Teacher's Role*, *Flexibility*, and *Adaptation*), and additional moderator variables, influence learning outcomes is worth investigating at a systematic review level as it falls under the overarching pursuit of a better understanding of what types of instructional environments ("which," "when," and "for what purpose") are most optimal for improving learning outcomes in a given instructional setting.

In this systematic review, the spectrum-based approach to K-12 T-C and S-C learning, and their inevitable combinations, by Bernard et al. (2019) will be applied to undergraduate level science post-secondary education in an attempt to delve further into the questions of "which," "when," and "for what purpose" regarding combinations of both learning approaches. Like in K-12, undergraduate level science post-secondary T-C and S-C learning approaches are not a dichotomy, but rather combinations as a function of the particular learning environment.

By applying the approach used by Bernard et al. (2019) to undergraduate level science post-secondary education, this systematic review hopes to answer the following research questions:

1) Does undergraduate level science post-secondary more S-C learning result in higher achievement outcomes (as measured by effect size) than undergraduate level science post-secondary less S-C (more T-C) learning, and does the degree of S-C (amount of S-C difference between treatment and control) predict the degree of achievement (amount of achievement outcome difference between treatment and control)?
2) Which primary predictor variables of student achievement (*Pacing, Teacher's Role, Flexibility, and Adaptation*) to what extent predict effect size, and what is the magnitude of effect as a function of the degree of S-C (amount of S-C difference between treatment and control) of each primary predictor variable?
3) What combinations of primary predictor variables of student achievement (*Pacing, Teacher's Role, Flexibility, and Adaptation*) better predict effect size?
4) Which moderator variables (technology use, subject matter, and treatment group class size) to what extent predict effect size, as well as which combinations of moderator variables predict effect size?

The outcomes of this meta-analysis will inform educational practitioners and the research community of the similarities and differences between T-C and S-C learning in K-12 and undergraduate level science post-secondary education, as well as what the more effective and less effective combinations are as a function of the specific undergraduate level science post-secondary learning environment.

Methods

Literature Search Strategy

Comprehensive literature searches were carried out by a fulltime Information Specialist (MLS level) and member of the Systematic Review Team at the Centre for the Study of Learning and Performance (CSLP) at Concordia University in Montréal, QC, Canada. The sources used for this systematic review were taken from a larger database that was created to explore S-C learning at various levels of education – from pre-kindergarten to post-secondary education. The same database was used for the literature search and retrieval process performed by Bernard et al. (2019) for their systematic review on S-C learning in K-12.

Inclusion/Exclusion Criteria

In order to be included in the meta-analysis, an individual study had to meet the following inclusion/exclusion requirements:

- Be publicly available (or archived) and encompass sources no earlier than 1960;
- Be conducted in formal undergraduate level post-secondary educational settings and address any of the following formal undergraduate level post-secondary education science subject matters: Biology, Chemistry, Physics, Geology or Psychology (Clinical and Experimental);
- Contain legitimate measures of academic achievement (i.e., instructor-made/researcher-made, standardized);
- Contain at least two groups of students receiving different instructional strategies/practices that can be compared as more S-C and less S-C instruction;

- Include course content and outcome measures that are compatible in the groups that form these comparisons;
- Contain sufficient descriptions of major instructional events that occurred in all instructional conditions;
- Fulfill requirements of either experimental or high-quality quasi-experimental design (QED); and
- Contain sufficient statistical information for effect size extraction.

As a result the studies included in the current meta-analysis could be characterized as follows.

Types of Studies

Only studies that considered the difference between two groups were eligible for inclusion. These studies were either experimental (i.e., RCTs) or high-quality QEDs (i.e., statistically verified group equivalence or adjustment) in design that adequately addressed the more S-C/less S-C group comparisons from the research questions, featured interventions that covered the same content (required knowledge acquisition and/or skill development) along with legitimate measures of academic achievement (i.e., instructor-made, standardized), and reported sufficient statistical information for effect size extraction.

Participants and Settings

The participants are undergraduate level science post-secondary students in formal educational settings eventually leading to a certificate, diploma, degree, or promotion to a higher level. The subject matter covered is science (biology, chemistry, physics, geology, and psychology). Educational interventions may take place either in the classroom/lecture hall (CI), via distance education (DE), or as a blended intervention (BL – various combinations of CI and DE).

Outcome Measures

All types of objective measures of academic achievements were considered. This included both standardized and non-standardized instructor/research-made assessment tools, as well as both cumulative final examinations and averages of several performance tasks covering various components of the course/unit content. Self- assessments were excluded, as well as attitudinal and behavioral measures. Data of their prevalence in the reviewed primary literature was collected to inform further reviews in the area with a potential focus on those types of outcomes.

Types of Interventions

The intervention in question (a treatment condition) was considered to be any combination of instructional events that is rated higher in S-C qualities than a comparison (control) condition. As such, the phenomenon being investigated in this meta-analysis is not an *intervention* in the typical way that this word is used in experimental literature. In this case, it is a set of instructional practices that have been rated along a continuum from extremely T-C to extremely S-C via scores on the four instructional events (i.e., dimensions) presented earlier, and used by, Bernard et al. (2019): *Pacing, Teacher's Role, Flexibility*, and *Adaptation*.

In this meta-analysis, teaching and learning have been deconstructed according to the events associated with them – a more S-C learning environment is one in which students play a more central role in the conduct of instructional events, and a more T-C learning environment is one in which instructional events are dominated by the instructors. As a result, any classroom research, regardless of the *intervention* being investigated, is eligible for inclusion as long as there is sufficient information provided as to what each participation group did.

Primary Predictor Variables

Two experienced reviewers working independently coded each participation group in each study from 1 (more T-C) to 5 (more S-C) on the same four primary predictor variables of student achievement (i.e., effect size-defining dimensions) presented earlier, and used by, Bernard et al. (2019): *Pacing, Teacher's Role, Flexibility*, and *Adaptation*. The reviewers had extensive experience from working on the coding for the Bernard et al. (2019) systematic review, as well as prior extensive training involving multiple practice runs on studies previously judged to have been accurately and reliably coded. Additionally, these same reviewers had both prior training and prior experience with the overall inclusion/exclusion criteria, and as a result were also assigned the task of determining which studies to include in this systematic review based on the requirements.

After the reviewers independently coded each participation group in each study (from 1 to 5 on each primary predictor variable), they met with a third member of the research team and proceeded to go over the coding. If there were any disagreements in coding (e.g., Reviewer #1 coded one participation group in one study as a 4 on *Pacing*, and Reviewer #2 coded that same participation group as a 2 on *Pacing*), the reviewers and the additional research team member would discuss the discrepancy until a consensus was reached. This procedure of independent coding followed by joint discussion and appropriate changes to the coding was employed at all stages of this systematic review.

Within each eligible comparative study, all participation groups were coded for the four effect size-defining dimensions using a five-point scale, as follows:

Pacing.

Encompasses the course design, as well as the selection and provision of study materials and the setting up of learning objectives.

- Degree to which instructors/students participate in various aspects of course planning (e.g., selection of study materials or setting learning objectives):
  1. No student involvement (most is determined by the instructor or program/curriculum)
  2. Student involvement in at least one of the components of course planning is present, but limited
  3. Instructors and students collaborate in the course planning, but instructor's role is still dominant
  4. Instructors and students collaborate in the course planning equally
  5. High student involvement – students play a leading role in course planning

Teacher's Role.

Represents a continuum of an instructor's major responsibilities for organizing/delivering instruction.

- Degree to which an instructor plays a predominant role in the teaching/learning process:

1. Instructor almost exclusively lectures, is the main source of content-relevant information and/or an authority figure
2. Instructor provides some guidance, feedback, initiates and supports discussions, etc.
3. Instructor functions as a guide, coach, tutor, provocateur of thinking
4. Instructor functions as a colleague, partner in learning
5. Instructor almost exclusively acts as a facilitator of learning, responding to students' specific needs (follows students' lead, consults, clarifies, encourages, etc.)

Flexibility.

Reflects the degree of student control over the time of instruction/learning (i.e., logistical flexibility) and over progression through the course content (i.e., pedagogical flexibility – revisiting/selecting/skipping/reordering topics and tasks).

- Degree to which students are given control over course progression:
  1. Highly structured instruction (no flexibility is allowed)
  2. Minor degree of either logistical or pedagogical flexibility is available to students
  3. Program/instructor's control over course progression is balanced with that of students
  4. Students have a substantial amount of flexibility in course progression
  5. High degree of flexibility (up to the point of completely self-paced and/or self-planned/self-managed learning)

Adaptation.

Describes the amount of modification in instructional process that is provided to accommodate individual students.

- Degree to which instruction takes into account students' needs/interests/level of knowledge:
  1. Learning materials, settings, activities and other work arrangements are predetermined and unchanged throughout the instruction (e.g., standardized or required curriculum)
  2. Minor modifications are allowed to either learning materials, group composition, or the context of instruction
  3. Elements of either individualized feedback, or role assignments and tasks, based on students' interests and/or previous achievements, etc.
  4. Adapting several instructional components (in combinations) to students' individual needs/interests/levels of knowledge
  5. High levels of joint adaptability of several components of instruction

Research Question 1

Does undergraduate level science post-secondary more S-C learning result in higher achievement outcomes (as measured by effect size) than undergraduate level science post-secondary less S-C (more T-C) learning, and does the degree of S-C (amount of S-C difference between treatment and control) predict the degree of achievement (amount of achievement outcome difference between treatment and control)?

Based on the results of the Primary Predictor Variable coding described above, numeric values were assigned to each participation group. The minimum score a group could receive was 4 (1 out of 5 on each of the four dimensions) and the maximum was 20 (5 out of 5 on each of the four dimensions). The sum of these values determined treatment (higher total out of 20 and therefore more S-C) and control (lower total out of 20 and therefore less S-C) conditions in every included study. While in any given participation group some dimensions might have been rated as more T-C (a score of either 1/5 or 2/5), more S-C (a score of 4/5 or 5/5), or equal (3/5), as long as one of the groups scored higher out of the total score of 20, a distinction between a more S-C group and a less S-C group was able to be made. Effect sizes of d-family (that is standardized mean differences) were then extracted from each individual study in order to measure the amount of difference in achievement outcomes between treatment and control (see "Extracting and Calculating Effect Sizes" section below). Afterwards, all effect sizes were averaged to determine the overall magnitude of impact of more S-C learning on undergraduate level science post-secondary student achievement outcomes.

A differential score between treatment and control was then calculated, with a range from 1 (one point difference in coding on only a single dimension, with the other three dimensions receiving equal scores for each group) to 16 (maximum difference between groups on all four coded dimensions, with one group receiving 1 out of 5 on each dimension, and the other group receiving 5 out of 5 on each dimension). Meta-regression analysis was then used to compare these differential scores with the student achievement outcomes reported for each participation group in order to explore whether degree of S-C predicted achievement outcomes across all studies.

Research Question 2

Which primary predictor variables of student achievement (Pacing, Teacher's Role, Flexibility, and Adaptation) to what extent predict effect size, and what is the magnitude of effect as a *function of the degree of S-C (amount of S-C difference between treatment and control) of each primary predictor variable?*

Based on the results of the Primary Predictor Variable coding described above, a differential score was calculated between the results of each participation group in a study on each of the four primary predictor variables (Treatment Group Score – Control Group Score = Differential Score). For each study, four differential scores were calculated – one for each dimension. In each study, a dimension was deemed more T-C if the differential score was between -1 and -4 (Treatment Group Score < Control Group Score), and was deemed more S-C if the differential score was between +1 and +4 (Treatment Group Score > Control Group Score). A score of 0 was interpreted as equality between conditions (Treatment Group Score = Control Group Score). While some participation groups might have scored more S-C on some dimensions and more T-C, or equal, on others, it is important to remember that the participation group with the higher total score out of 20 was deemed the treatment (more S-C) condition in each study. Meta-regression analysis was then used to compare the differential scores of each dimension with the student achievement outcomes reported for each participation group in order to explore the relationship between degree of S-C in each dimension and achievement outcomes (with achievement outcomes being the outcome variable and defined as the average effect size of all the studies).

Mixed moderator variable analysis was then used to explore the degree of S-C (amount of S-C difference between treatment and control) of each of the four primary predictor variables and their individual magnitude of effect on post-secondary student achievement outcomes. For each dimension, the degree (level) of S-C ranged from 0 (S-C = T-C) to 4 (S-C >>>> T-C). For each degree of S-C being examined, achievement outcomes were the outcome variable and were defined as the average effect size of all the studies that contained the same degree of S-C on the dimension in question (e.g., When examining the dimension of Flexibility at level 2 (S-C >> T-C), only studies that contained level 2 Flexibility were included in the analysis).

Research Question 3

What combinations of primary predictor variables of student achievement (Pacing, Teacher's Role, Flexibility, and Adaptation) better predict effect size?

Meta-regression analysis was used to compare primary predictor variables alone (i.e., *Pacing; Teacher's Role*) with the primary predictor variables paired (e.g. *Pacing + Teacher's Role*).

Instructional and Demographic Moderator Variables

The same two experienced reviewers who independently coded the primary predictor variables also coded the following three moderator variables: Technology Use (instructional moderator variable), Subject Matter (demographic moderator variable), and Treatment Group Class Size (demographic moderator variable). Technology Use was coded as either Yes or No; Subject Matter as either Biology, Chemistry, Physics, Geology or Psychology; and Treatment Group Class Size as either Small (15 students or under), Medium (16 to 49 students), Large (50 to 99 students) or Very Large (100+ students), regardless what the class size of the control group was.

Research Question 4

Which moderator variables (technology use, subject matter, and treatment group class size) to what extent predict effect size, as well as which combinations of moderator variables predict effect size?

Following the moderator variable coding, mixed moderator variable analysis and multiple meta-regression analysis were used to explore which variables on their own, as well as which combinations of variables, influenced student learning achievement outcomes.

Extracting and Calculating Effect Sizes

In order for studies to contain sufficient information for effect size extraction, the following statistical information was considered (in all cases sample size data were required):

- Means and standard deviations for each treatment and control group;
- Exact t-value, F-value, with an indication of the ± direction of the effect;
- Exact p-value (e.g., $p = .012$), with an indication of the ± direction of the effect;
- Effect sizes converted from correlations or log odds ratios;
- Estimates of the mean difference (e.g., adjusted means, regression □ weight, gain score means when r is unknown);
- Estimates of the pooled standard deviation (e.g., gain score standard deviation, one-way ANOVA with three or more groups);
- Estimates based on a probability of a significant t-test using □ (e.g., $p < .05$); and
- Approximations based on dichotomous data (e.g., percentages of students who succeeded or failed the course requirements).

Effect sizes were initially calculated as Cohen's *d* and then converted to Hedges' *g* (i.e., correction for small samples). Standard errors ($SE_{\bar{d}}$) were calculated for $\bar{d}$ and then converted to standard errors of $SE_{\bar{g}}$ applying the correction formula for *g*. Hedges' *g*, $SE_{\bar{g}}$ and sample sizes (i.e., treatment and control) were entered into *Comprehensive Meta-Analysis 3.3.07* (Borenstein, Hedges, Higgins, & Rothstein, 2014) where statistical analyses were performed.

In all of these analyses, including the multiple regression, the inverse-weighted random effects model was used for interpretation. Average effect sizes, therefore, are symbolized as $\bar{g}_{Random}$. Under the tenets of this model, as contrasted with the fixed effect model, between-study variation is not collected and analyzed separately (i.e., Q_{Total}), but instead is incorporated into each inverse-variance weighted (*W*) effect size as τ^2 (tau-squared) that make up $\bar{g}_{Random}$, thus, $W_{Random} = 1 / v_{within} + \tau^2$. This tends to make the overall average effect size ($\bar{g}_{Random}$) more conservative and the overall standard ($SE_{\bar{g}}$) larger, and also more conservative. Other overall statistics, such as the significance tests *z* and *p*, also tend to be more conservative.

The effect sizes were coded for precision of calculations and analyzed in subsequent moderator variable analysis.

Results

The Results section begins with a description of the literature search and retrieval process, followed by an analysis of bias. The subsequent synthesis of results provides results for each of the four research questions. The moderator variable analysis for research question #4 begins with categorical-level analysis, followed by multiple meta-regression analysis.

Description of Studies

Results of the search.

Literature search and retrieval was performed by a fulltime Information Specialist (MLS level) and member of the Systematic Review Team at the Centre for the Study of Learning and Performance at Concordia University in Montréal, QC, Canada. The sources used for this systematic review were taken from a larger database that was created to explore S-C learning at various levels of education – from pre-kindergarten to post-secondary education. The same database was used for the literature search and retrieval process performed by Bernard et al. (2019) for their systematic review on S-C learning in K-12.

The database as a whole initially contained 9759 total search results obtained from searching through 14 different databases and using both the Google and Bing search engines. Duplicate sources were removed (1285) and the remaining 8474 sources were subjected to an abstract screening process, which yielded 3749 sources for subsequent full-text retrieval and review. Of those, 167 were not retrievable, 2751 were excluded, and the remaining 831 were included. Examination of these 831 full-text sources proceeded according to the details described in the

Methods section. A total of 735 sources were excluded due to the inclusion/exclusion criteria in the Methods section, results in 96 included sources. Sources containing Technology, Engineering and Mathematics as the subject matter were also part of the 735 excluded sources – the 96 included sources were all Science (S of STEM included; TEM of STEM excluded). The final stage consisted of the extraction, coding, and analysis of 141 independent effect sizes from these 96 included sources. References to these 96 individual sources appear in the section *References to included sources*. Statistical information for the 141 independent effect sizes appears in Appendix 1 – *Descriptive Statistics for Each Study*.

Bias Analysis: Research Design, Publication and Sensitivity

Regarding research design, only quasi-experimental (QED) and random assignment (RCT) experiments were included (Table 1), with QED producing a slightly higher moderate effect size ($\overline{g}$ = 0.40 vs. 0.25) from a larger pool of studies (k = 85 vs. 56). The model produced a statistically non-significant result (Q-Between = 3.15, df = 1, p = 0.08).

Table 1
Research design bias analysis

Codes	k	$\overline{g}$	SE	Lower 95th	Upper 95th	z-value	p-value	Q-B	df	p-value
QED	85	0.40	0.05	0.30	0.50	7.94	< .001			
RCT	56	0.25	0.07	0.12	0.38	3.68	< .001			
Total between								3.15	1	.08

Publication date was subjected to an analysis containing five options (Table 2): 1960-1979; 1980-1989; 1990-1999; 2000-2009; and 2010-2012. 1960-1979 and 2010-2012 both produced non-significant results (*p* = 0.45 and *p* = 0.12). 2000-2009 produced a statistically significant moderate effect size ($\overline{g}$ = 0.34) from the largest sample of effect sizes (*k* = 81 out of 141). While 1980-1989 produced the highest effect size ($\overline{g}$ = 0.65), it only contained *k* = 12 out of the possible 141 effect sizes. The overall model produced a statistically significant result (*Q*-Between = 20.46, *df* = 4, *p* < .001).

Table 2
Publication date analysis

Levels	*k*	$\overline{g}$	*SE*	Lower 95th	Upper 95th	*z*-value	*p*-value	*Q*-Bet.	*df*	*p*-value
1960-1979	9	0.07	0.07	-0.11	0.25	0.75	.45			
1980-1989	12	0.65	0.65	0.28	1.01	3.44	.001			
1990-1999	20	0.16	0.16	-0.00	0.32	1.96	.05			
2000-2009	81	0.34	0.43	0.52	0.52	8.72	< .001			
2010-2012	19	0.15	0.12	-0.09	0.39	1.24	.12			
Total between								20.46	4	< .001

Analysis of publication bias seeks to determine if a sizable number of studies might have been missed or otherwise not included in a meta-analysis (Rothstein et al., 2005) and that this number, if found and included, would nullify the average effect. Publication bias was explored via a Funnel Plot, Duval and Tweedies's Trim and Fill (2000) procedure, as well as Classic and Orwin's fail-safe *N* procedures.

The Funnel Plot (Figure 1) indicates that there was no discernable publication bias on the negative side of the plot (i.e., left of the mean effect size). The Trim and Fill results suggest a similar pattern of inclusiveness. Classic fail-safe *N* suggests that 9749 additional effect sizes would be needed to bring the observed *p*-value below alpha = .05 (i.e., 69.1 missing studies would be needed for every observed study for the effect to be nullified). Lastly, Orwin's fail-safe *N* (Orwin, 1983) suggests that 113 additional 'null' effect sizes would be needed to bring the observed average effect size to a trivial level of $\bar{g}$ = 0.15.

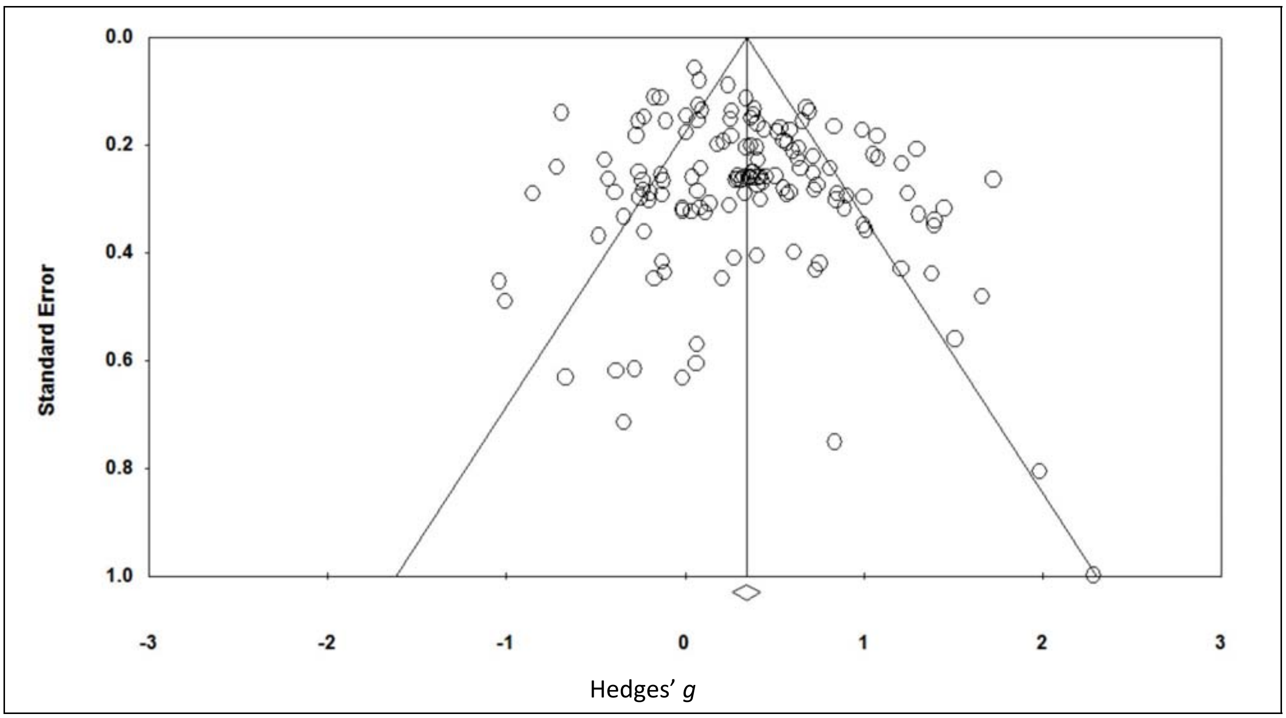

Figure 1. Funnel plot of 141 effect sizes (Hedges' *g* – X-Axis; Standard error – Y-Axis).

The final analysis of bias, sensitivity analysis, seeks to determine if effect sizes, especially at the upper and lower ends of the distribution (where higher and lower effect sizes are sometimes paired with anomalously large sample sizes) have any undue influence on the overall random effects outcomes (Borenstein et al., 2009). Table 3 shows the six highest and six lowest effect sizes and the overall influence when they are systematically removed from the distribution and the results effects are recalculated (i.e., one study removed, as exemplified below). The overall procedure was conducted using the software Comprehensive Meta-Analysis (Borenstein et al., 2014).

Column 1 is the Study Name and Date of Publication. Column 2 is the actual calculated *g* for each of the six highest and six lowest effect sizes. Columns 3 through 8 are the recalculated statistics when each study is removed and the statistics recalculated. Column 9 is the relative weight that is applied under the random model. Higher weights produce more influence than lower weights.

There appear to be no anomalous results across the 12 studies. This suggests that there is little or no 'effect size by sample size bias', at least at the extremities. This does not mean that there is no bias within the remaining 129 studies, but it is likely that if bias is present in these smaller effect size studies the overall results will not be as affected as it would in these 12 studies. As a result of this analysis no effect sizes were removed as outliers and no study was Winsorized (i.e., given the value of the next highest or lowest study).

Table 3
Sensitivity analysis (random effects)

Study Names	Actual g	One Study Removed						Relative Weight
		$\bar{g}$	*SE*	Lower 95th	Upper 95th	z-Value	p-value	
Wozniak2012	2.30	0.34	0.04	0.26	0.42	8.53	0.00	0.14
Okebukola1988	1.99	0.34	0.04	0.26	0.37	8.51	0.00	0.20
Doymus2008	1.73	0.33	0.03	0.25	0.35	8.46	0.00	0.73
Okebukola1988	1.66	0.34	0.04	0.26	0.37	8.47	0.00	0.34
Folconer1988	1.51	0.34	0.04	0.26	0.38	8.49	0.00	0.64

Swanson1990	-0.48	0.35	0.04	0.27	0.42	8.68	0.00	0.29
Kapp2011	-0.69	0.35	0.04	0.28	0.42	9.04	0.00	0.94
Hulshof2005	-0.72	0.35	0.04	0.27	0.43	8.84	0.00	0.77
Reinhardt2012	-0.85	0.35	0.04	0.27	0.43	8.84	0.00	0.68
Martin2009	-1.01	0.35	0.04	0.27	0.43	8.73	0.00	0.41
Martin2009	-1.04	0.35	0.04	0.27	0.43	8.76	0.00	0.45
Overall (k = 141)	0.34	—	0.04	0.26	0.42	8.58	0.00	100.00

Synthesis of Results

Research question 1.

Does undergraduate level science post-secondary more S-C learning result in higher achievement outcomes (as measured by effect size) than undergraduate level science post-secondary less S-C (more T-C) learning, and does the degree of S-C (amount of S-C difference between treatment and control) predict the degree of achievement (amount of achievement outcome difference between treatment and control)?

While the first research question considered the overall average effect on achievement outcomes of more adaptive science instruction as it is reflected in the difference between more S-C learning (treatment condition) and less S-C learning (control condition), it is important to remember that it is not necessarily a mutually exclusive difference between S-C learning environments and T-C environments. Rather, as outlined in the Methods section, it is a difference in ratings on four effect size-defining dimensions (*Pacing, Teacher's Role, Flexibility*, and *Adaptation*) between each participation group, with the group receiving a higher total score being the treatment condition, and the group receiving a lower total score being the control condition.

One hundred and forty one unadjusted effect sizes were included in the meta-analysis (Table 4), producing a significant random effects model weighted average effect size of $\overline{g}$ = 0.34, k = 141, SE = 0.04, z = 8.58, p < .001. The distribution was significantly heterogeneous (Q-Total = 618.13, df = 140, p < .001, with an I^2 value of 0.77 and a tau-squared (τ^2) of 0.15), suggesting that a large degree of between-study variance was present. The results of this analysis suggest that on average *more* S-C classroom studies produce better results on achievement outcomes than do *less* S-C classroom studies. The average weighted effect size is of moderate size (Cohen, 1988) and indicates that on average the more S-C condition (treatment) outperformed the less S-C condition (control) by 0.34*sd*.

Table 4
Overall results

Model	Effect size and 95th Confidence Interval					Test of null	
Random Effects	k	$\overline{g}$	SE	Lower 95th	Upper 95th	z-value	p-value
Total Collection	141	0.34	0.04	0.26	0.42	8.58	< .001
Model	Between-group Heterogeneity						
Fixed Effect	Q-value	df	p-value	I^2	Tau^2		
Total Collection	618.13	140	< .001	0.77	0.15		

Meta-regression analysis can further provide a sense of the relationship between the degree of S-C and achievement outcomes. As outlined in the Methods section, a measure of the degree of S-C (the quantitative differences between the ratings of the treatment/control) was calculated to test this relationship. If this relationship is patterned (either positively or negatively) rather than irregular, the result of the meta-regression of achievement on the degree of student-centeredness should result in a positive, significant slope. If the slope of the regression line is not positive and significant, indicating the absence of a positive linear progression, we can assume that the relationship between student-centeredness and achievement results is irregular, thereby diminishing the argument that more S-C classrooms are more advantageous to attaining higher achievement outcomes than less S-C classrooms.

The simple meta-regression of the relative difference between more S-C and less S-C (Table 5) resulted in a non-significant slope ($\beta = 0.01$, $SE = 0.01$, $z = 0.03$, $p = 0.40$). The test of the model resulted in Q-Between = 0.70, $df = 1$, $p = 0.40$, which is also non-significant. These results indicate an extremely weak, non-significant relationship between the degree of S-C and student learning achievement outcomes.

Table 5
Simple meta-regression: Overall strength of the relationship between treatment and control (degree of student-centeredness)

Covariate	β	*SE*	Lower 95th	Upper 95	*z*-value	*p*-value	VIF
Intercept	0.28	0.08	0.12	0.44	3.43	< .001	4.23
Total difference (across 4 predictors)	0.01	0.01	-0.01	0.03	0.03	.40	1.00
Test of Model: $Q_{Between} = 0.70$, $df = 1$, $p = .40$							

Note: This and all subsequent meta-regression analyses use Random Effects Method-of-Moments Model.

Research question 2.

Which primary predictor variables of student achievement (Pacing, Teacher's Role, Flexibility, and Adaptation) to what extent predict effect size, and what is the magnitude of effect as a function of the degree of S-C (amount of S-C difference between treatment and control) of each primary predictor variable?

As outlined in the Methods section, a differential score (Treatment Group Score – Control Group Score = Differential Score) was calculated between the results of each participation group in a study on each of the four dimensions of classroom practice (primary predictor variables). A differential score between -1 and -4 meant a dimension was more teacher-centered, and a score between +1 and +4 meant a dimension was more student-centered. A score of 0 meant equality. The four dimensions (described in detail in the Methods section) are:

- *Teacher's Role* as a lecturer/guide/mentor;
- *Pacing* of instruction to meet student needs/preferences;
- *Adaptability* of feedback and learning activities to students, individual interests of students, etc.
- *Flexibility* in the creation/use of study materials, course design, etc.;

Meta-regression analysis was used to explore which, if any, of these dimensions predict effect size (student achievement outcomes). All four dimensions were entered into multiple meta-regression (random effects method of moments) in the order that they are described above. The dependent, or outcome, variable in this analysis was the average effect size of all the studies (i.e., the student achievement outcomes).

The overall model, excluding the intercept (Table 6), was non-significant (*Q*-Between = 6.47, *df* = 4, *p* = 0.17). Additionally, none of the four dimensions were significant predictors of effect size, with *Teacher's Role* being the closest to reaching the significance level (β = 0.08, *SE* = 0.06, *z* = 1.50, *p* = 0.13). It should also be noted that *Teacher's Role* was the only dimension to have a positive (non-significant) relationship with effect size (student achievement outcomes). The other three dimensions – *Pacing* (β = -0.03, *SE* = 0.05, *z* = -0.58, *p* = 0.56); *Adaptability* (β = -0.41, *SE* = 0.07, *z* = -0.75, *p* = 0.45); and *Flexibility* (β = -0.09, *SE* = 0.05, *z* = -1.87, *p* = 0.61) – all had negative (non-significant) relationships with effect size.

Table 6
Meta-regression analysis – All four predictors

Covariates	β	*SE*	Lower 95th	Upper 95	*z*-value	*p*-value	VIF
Intercept	0.34	0.08	0.18	0.51	4.20	< .001	4.196
Teacher's Role	0.08	0.06	-0.03	0.19	1.50	.13	1.06
Pacing	-0.03	0.05	-0.03	0.19	-0.58	.56	1.06
Adaptation	-0.41	0.07	-0.18	0.08	-0.75	.45	1.04
Flexibility	-0.09	0.05	-0.19	0.00	-1.87	.61	1.04
Test of Model: Q_{Between} = 6.47. *df* = 4, *p* = .17							

The analysis was re-run with only *Flexibility* (Table 7) and this model produced a significant result (*Q*-Between = 3.80, *df* = 1, *p* ≤ ..05). As in the previous analysis, *Flexibility* also had a negative relationship with a negative slope (β = -0.09).

Table 7
Simple meta-regression: Flexibility alone

Covariate	β	*SE*	Lower 95th	Upper 95	*z*-value	*p*-value	VIF
Intercept	0.39	0.05	0.30	0.48	8.38	< .001	1.36
Flexibility	-0.09	0.05	-0.19	0.00	-1.95	.05	1.00
Test of Model: $Q_{\text{Between}} = 3.80$, $df = 1$, $p = .05$							

Regarding the degree of S-C of each of the four dimensions and their individual magnitude of effect on student achievement outcomes, mixed moderator variable analysis was used to explore the effect that each degree of S-C for each dimension had on the average effect size of all the studies that contained the same degree of S-C on the dimension in question.

Of the four dimensions, only *Flexibility* (Table 8) was significant (Q-Between = 9.30, $df = 4$, $p = 0.05$). *Pacing* (Q-Between = 0.28, $df = 3$, $p = 0.96$), *Teacher's Role* (Q-Between = 0.96, $df = 3$, $p = 0.81$), and *Adaptability* (Q-Between = 0.23, $df = 1$, $p = 0.63$) were all far from significant, thus indicating no significant relationship between the degree of S-C of each of these three dimensions and student achievement outcomes. While *Flexibility* was the only significant dimension, it should be noted that it had a negative relationship between degree of S-C and effect size. At differential scores of 0 ($\overline{g} = 0.37$), +1 ($\overline{g} = 0.33$) and +2 ($\overline{g} = 0.35$), *Flexibility* produced positive effect sizes of moderate size, however with higher degrees of *Flexibility*, a negative relationship was present ($\overline{g} = -0.44$ at differential score = +3, and $\overline{g} = -0.19$ at differential score = +4).

Table 8
Levels of flexibility across levels of hedges' $\overline{g}$

Levels	*k*	$\overline{g}$	*SE*	Lower 95th	Upper 95th	*z*-value	*p*-value	*Q*-Bet.	*df*	*p*-value

0 – no difference	95	0.37	0.05	0.27	0.48	7.00	< .001			
+1 – Favors SC over TC	31	0.33	0.07	0.21	0.46	5.10	< .001			
+2 – Favors SC over TC	10	0.35	0.11	0.13	0.56	3.17	.002			
+3 – Favors SC over TC	3	-0.44	0.49	-1.40	0.51	-0.91	.37			
+4 – Favors SC over TC	2	-0.19	0.21	-0.60	0.22	-0.91	.37			
Between groups								9.30	4	.05

Course Pacing (*Q*-Between = 0.28, *df* = 3, *p* = .96)
Teacher's Role (*Q*-Between = 0.96, *df* = 3, *p* = .81)
Adaptation of Materials and Methods (*Q*-Between = 0.23, *df* = 1, *p* = .63)

Research question 3.

What combinations of primary predictor variables of student achievement (Pacing, Teacher's Role, Flexibility, and Adaptation) better predict effect size?

Meta-regression analysis was used to compare primary predictor variables alone (i.e., *Pacing; Teacher's Role*) with the primary predictor variables paired (e.g. *Pacing* + *Teacher's Role*), however there were no substantial, significant findings. This lends to the conclusion that combinations of instructional dimensions are not indicative of influence on student achievement outcomes.

Research question 4.

Which moderator variables (technology use, subject matter, and treatment group class size) to what extent predict effect size, as well as which combinations of moderator variables predict effect size?

Technology use (Table 9) was coded as either "yes" or "no" and three options were subjected to analysis: treatment = yes vs. control = yes (1_1); treatment = yes vs. control = no (1_2); and treatment = no vs. control = no (2_2). Theoretically a fourth option of treatment = no vs. control = yes (2_1) was possible, but it was not present in this data set. All three options produced statistically significant effect sizes, however 1_1 produced an effect size of small size ($\overline{g}$ = 0.19) compared to the two moderate effect sizes produced by 1_2 ($\overline{g}$ = 0.48) and 2_2 ($\overline{g}$ = 0.40). The model itself was also statistically significant (*Q*-Between = 9.06, *df* = 2, *p* = 0.01).

Table 9
Technology use analysis

Levels	*k*	$\overline{g}$	*SE*	Lower 95th	Upper 95th	*z*-value	*p*-value	*Q*-B	*df*	*p*-value
Yes_Yes	51	0.19	0.06	0.00	0.30	3.55	.001			
Yes_No	23	0.48	0.11	0.01	0.70	4.10	< .001			
No_No	67	0.40	0.06	0.28	0.52	6.865	< .001			
Total between								9.06	2	.01

Post hoc difference between No-No[1] & Yes-Yes[2] (*Q*-B = 6.39, df = 1, *p* = .01).
Post hoc difference between Yes-Yes[1] & Yes-No[3] (*Q*-B = 6.39, df = 1, *p* = .01).

Technology use was further analyzed across degrees of *Flexibility* (Table 10). Each of the three codes from Table 9 (2_2; 1_1; and 1_2) was analyzed at *Flexibility* differential scores of 0

(S-C = T-C), 1 (S-C > T-C), and 2+ (S-C >> T-C; S-C >>> T-C; and S-C >>>> T-C). Of the three technology use codes, only 2_2 (neither group used technology by degrees of *Flexibility*) produced a significant model (Q-Between = 6.93, df = 2, p = 0.03). Furthermore, 2_2 also produced a significant declining effect size average over levels of *Flexibility*. While both 1_1 (both groups used technology by degrees of *Flexibility*) and 1_2 (treatment group used; control group didn't by degrees of *Flexibility*) were non-significant, 1_1 also had a negative relationship between degree of *Flexibility* and effect size, whereas 1_2 had a positive relationship.

Table 10
Levels of technology use by levels of flexibility (Codes 0, 1 and 2+)

Levels	k	$\bar{g}$	SE	Lower 95th	Upper 95th	z-value	p-value	Q-Bet.	df	p-value
a) No-No (Neither group used technology by levels of flexibility – Codes 0, 1 and 2+)										
Both groups "No" & 0	39	0.52	0.08	0.35	0.70	6.20	< .001			
Both groups "No" & 1	16	0.34	0.10	0.14	0.54	3.26	.001			
Both groups "No" & 2+	12	0.11	0.06	-0.154	0.70	.86	< .39			
Between groups								6.93	2	.03
b) Yes-Yes (Both groups used technology by levels of flexibility – Codes 0 and 1)										
Both groups "Yes" & 0	41	0.20	0.07	0.07	0.33	3.09	.001			
Both groups "Yes" & 1	9	0.13	0.09	-0.04	0.31	1.49	.14			
Between groups								.39	1	.53
c) Yes-No (Treatment group used; control group didn't by levels of flexibility – Codes 0 and 1)										
Treat. "Yes," Cont. "No" & 0	15	0.44	0.18	0.15	0.31	2.56	.01			
Treat. "Yes," Cont. "No" & 1	8	0.53	0.11	0.31	0.74	4.82	< .00			
Between groups								.14	1	.71

Codes for flexibility: 0 means T-C and S-C are equal; 1 means S-C > T-C; 2+ means S-C >> T-C

The subject matter covered in this meta-analysis was science-based (Table 11), with the subject breakdown consisting of: biology, chemistry, geology, physics and psychology (clinical and experimental). The overall model produced a statistically significant result (*Q*-Between = 20.93, $df = 4$, $p < .001$). Geology ($\bar{g} = -0.05$) and psychology ($\bar{g} = 0.06$) were the only subjects to yield non-significant results ($p = 0.85$ and $p = 0.51$, respectively). Biology ($\bar{g} = 0.31$) and physics ($\bar{g} = 0.45$) produced moderate effect sizes, and chemistry ($\bar{g} = 0.58$) produced a large effect size. Of the five subjects explored, geology was the only one with a negative effect size ($\bar{g} = -0.05$).

Table 11
Subject matter analysis

Levels	*k*	$\bar{g}$	*SE*	Lower 95th	Upper 95th	*z*-value	*p*-value	*Q*-Bet.	*df*	*p*-value
Biology	46	0.31	0.06	0.20	0.42	5.56	< .001			
Chemistry	24	0.58	0.01	0.40	0.76	6.26	< .001			
Geology	5	-0.05	0.26	0.07	-0.55	0.45	.85			
Physics	37	0.45	0.01	0.28	0.61	5.27	< .001			
Psychology	20	0.06	0.09	-0.12	0.24	0.65	.51			
Total between								20.93	4	< .001

Treatment group class size (Table 12) options were: Small (15 students or under); Medium (16 to 49 students); Large (50 to 99 students); and Very Large (100+ students). The overall model was significant (*Q*-Between = 15.51, $df = 3$, $p = 0.001$). Of the four options, only Medium ($p < 0.001$) and Large ($p < 0.000$) were significant, as well as both produced the two highest effect sizes ($\bar{g} = 0.40$ and $\bar{g} = 0.46$, respectively).

Table 12
Class size of treatment group analysis

Levels	k	$\overline{g}$	SE	Lower 95th	Upper 95th	z-value	p-value	Q-Bet.	df	p-value
Small	27	0.29	0.16	-0.01	0.61	1.90	.06			
Medium	76	0.40	0.06	0.28	0.51	6.86	< .001			
Large	20	0.46	0.08	0.01	0.62	5.53	< .000			
Very Large	18	0.09	0.07	-0.05	-0.23	1.20	.23			
Total between								15.51	3	.001

Note: Small = ≤ 15 students; Medium = 16 to 49 students; Large = 50 to 99 students; Very Large = 100+ students.

The above categorical coding produced a number of significant moderator variables, which were subsequently subjected to multiple meta-regression analysis to further provide a sense of the relationship between the moderator variables and student achievement outcomes. The first multiple meta-regression (Table 13) examined the two integer-level variables of treatment group class size and flexibility. Both class size (β = -0.001, SE = 0.0001, z = -1.93, p = 0.05) and flexibility (β = -0.09, SE = 0.05, z = -2.01, p = 0.05) had a significant, negative relationship with effect size – as class size gets smaller average student achievement outcomes go up (Figure 2), and as flexibility goes up (greater S-C learning) average student achievement outcomes go down (Figure 3).

Table 13
Multiple meta-regression: Flexibility and treatment group class size (Integer-Level)

Covariate (Predictors)	β	SE	Lower 95th	Upper 95th	z-value	p-value	VIF
Intercept	0.45	0.06	0.34	0.56	8.03	< .001	2.00
Class Size (*Treatment)	-0.001	0.0001	-0.002	-0.000	-1.93	.05	1.00
Flexibility	-0.09	0.05	-0.19	0.002	-2.01	.05	1.00

Test of Model: *Q*-Between (predictors) = 7.52, *df* = 2, *p* = .02; R^2 = 0.02 or 2%
Homogeneity: *Q*-Between (studies) = 567.82, *df* = 138, *p* < .0001, I^2 = 75.70, tau^2 = 0.1444

*Class sizes were taken from treatment groups of 141 effect sizes.

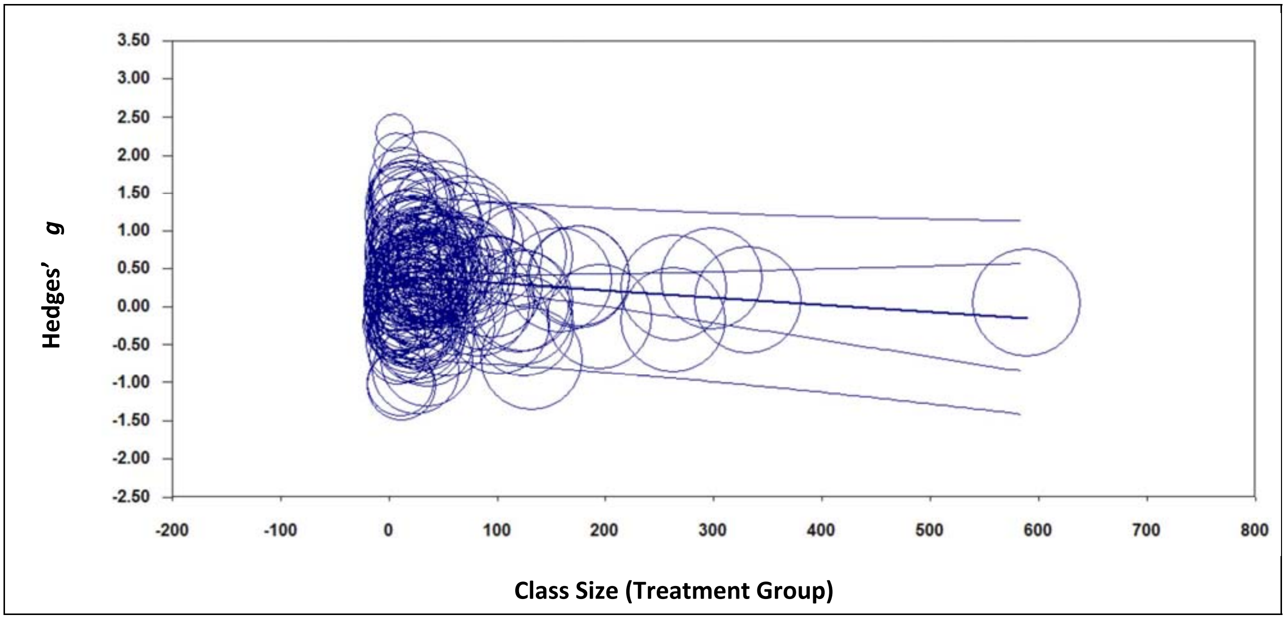

Figure 2. Scatterplot showing Hedges' *g* by class size.

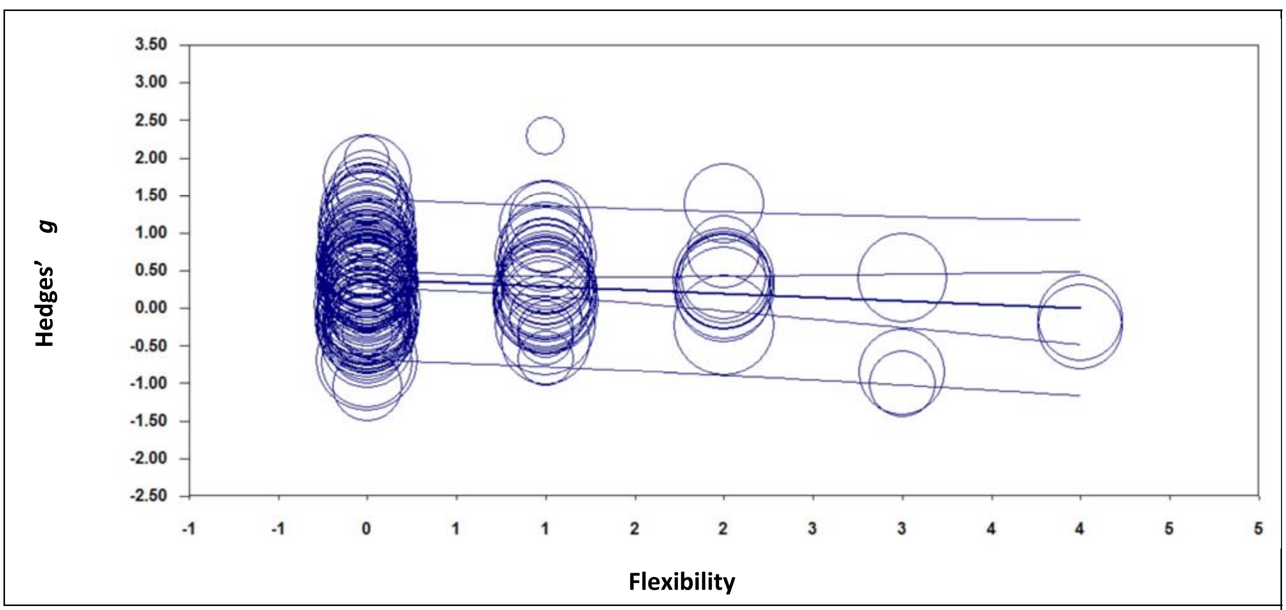

Figure 3. Scatterplot showing Hedges' *g* by flexibility.

Multiple meta-regression analysis was also performed for the four predictors of treatment group class size, flexibility, subject matter, and technology use (Table 14). Along with the integer-level predictors, the overall models for subject matter (Q-Between = 20.71, df = 4, p < 0.001) and technology use (Q-Between = 6.47, df = 4, p = 0.04) were also significant. The overall model (Q-Between (predictors) = 42.24, df = 8, p = < .0001, R^2 = 0.26 or 26%; and Q-Between (studies) = 430.65, df = 132, p = < .0001, I^2 = 69.35%, tau^2 = 0.011) was also significant. It is also important to note that R^2 = 0.26, indicating that 26% of the variance is accounted for. As in Table 13, both class size and flexibility were negative predictors of effect size. Regarding subject matter (Figure 4), chemistry was the best predictor of effect size (β = 0.23, SE = 0.11, z = 2.13, p = 0.03), with a significantly higher average effect size than the reference group of biology. The others (except physics, which is not significant) were negative, meaning that they had significantly lower average effect sizes than biology. Regarding technology use (Figure 5), "both groups yes" (BGY) had a significantly lower average effect size (β = -0.18, SE = 0.09, z = -2.09, p = 0.04) than the reference group of "both groups no" (BGN). When the treatment group has technology and the control group doesn't (EYCN), the result was not significantly different from BGN (β = 0.07, SE = 0.11, z = 0.06, p = 0.53). In other words, the presence of technology in the treatment group does not outperform studies where no technology is in either group.

Table 14
Multiple meta-regression of four predictors: Treatment group class size, flexibility, subject matter, and technology use

Covariates (Predictors)	β	*SE*	Lower 95th	Upper 95th	*z*-value	*p*-value	VIF
Intercept	0.51	0.08	0.34	0.67	6.13	< .00	5.25
			Integer Predictors				
Class Size (Treatment)	-0.001	0.00	-0.002	-0.0001	-2.24	.03	1.02
Flexibility	-0.13	0.05	-0.22	-0.05	-2.83	.005	1.14
	[1]Subject Matter (Categorical) $Q = 20.71$, $df = 4$, $p < .0001$						
Chemistry	0.23	0.11	-0.18	0.44	2.13	.03	1.34
Geology	-0.45	0.21	-0.86	-0.04	-2.14	.03	1.60
Physics	0.10	0.098	-0.01	0.29	0.98	.33	1.38
Psychology	-0.25	0.11	-0.47	-0.03	-2.24	.02	1.19
	[2]Technology Use (Categorical) $Q = 6.47$, $df = 2$, $p = .039$						
Yes-Yes	-0.18	0.09	-0.35	-0.11	-2.09	.04	1.34
Yes-No	0.07	0.11	-0.14	0.28	0.06	.53	1.37

Test of Model: *Q*-Between (predictors) = 42.24, $df = 8$, $p = < .0001$, $R^2 = 0.26$ or 26%; (Between predictors: variance accounted for by the model).

Homogeneity: *Q*-Between (studies) = 430.65, $df = 132$, $p = < .0001$, $I^2 = 69.35\%$, $tau^2 = 0.011$ (Between Studies: Variance not accounted for).

[1] Reference Group for Subject Matter: Biology.
[2] Reference Group for Technology Use: Both groups used no technology (No-No).

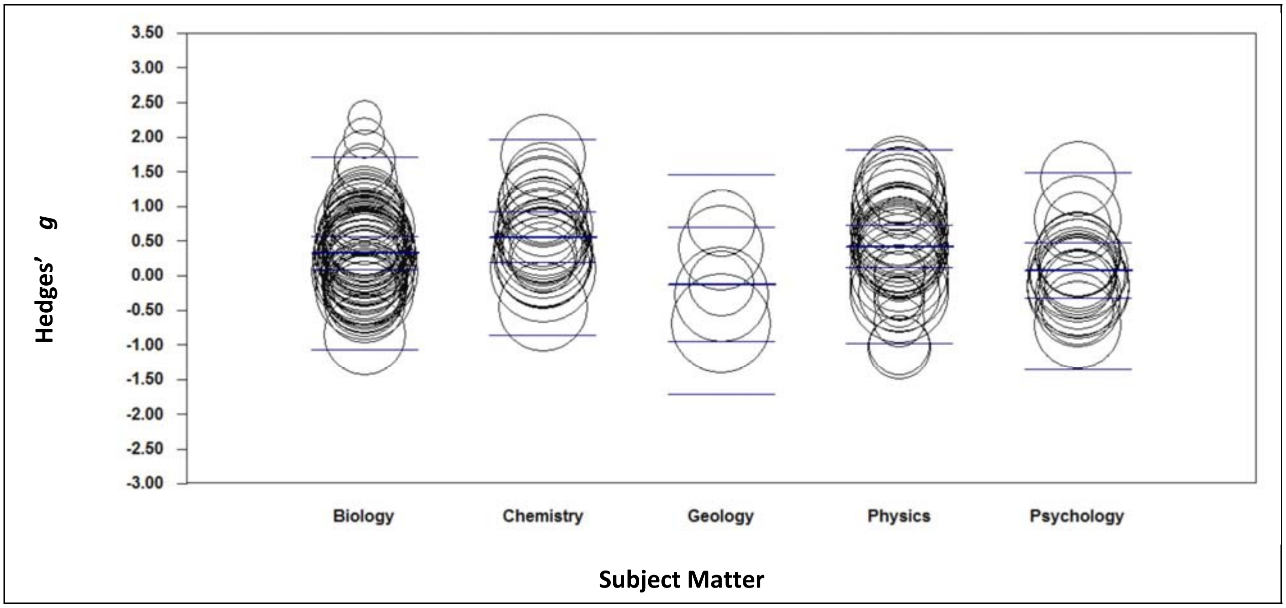

Figure 4. Scatterplot showing Hedges' *g* by subject matter.

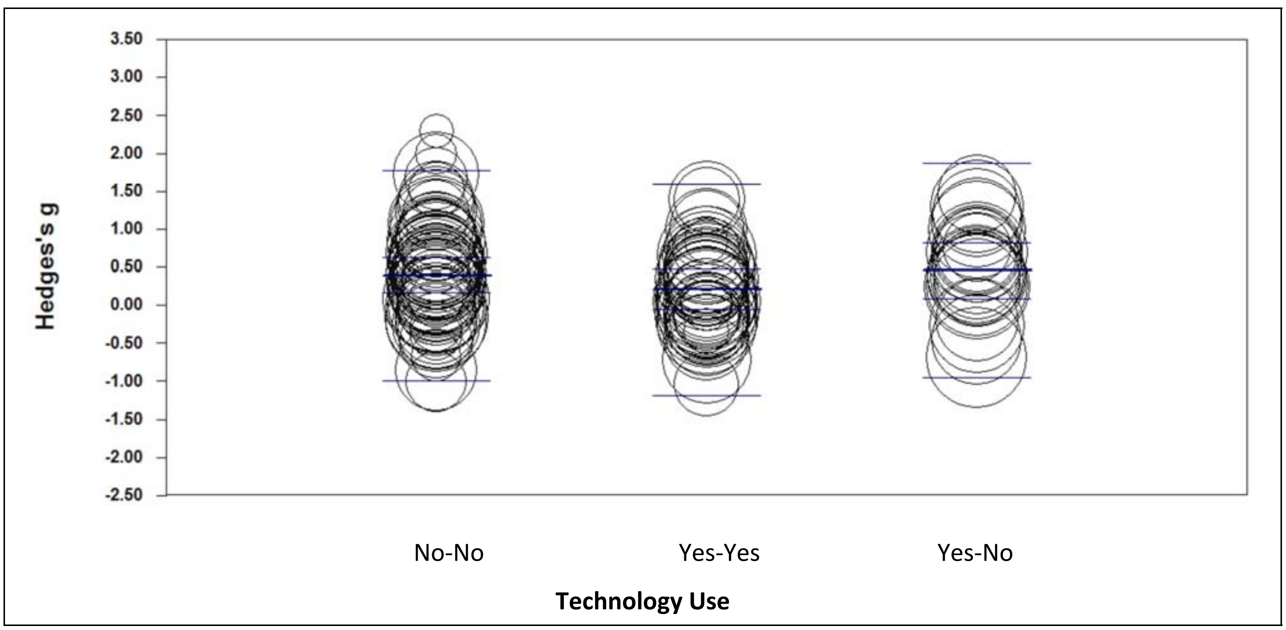

Figure 5. Scatterplot showing Hedges' *g* and technology use.

Codes: No-No (both groups had no technology); Yes-Yes (both groups had technology); and Yes-No (treatment group had technology, control Group didn't).

Discussion

The purpose of this systematic review is to examine the effectiveness of S-C instructional practices in undergraduate level science post-secondary settings as it pertains to student achievement outcomes. It further seeks to examine the individual/collective influence of four instructional dimensions (*Pacing, Teacher's Role, Flexibility,* and *Adaptation*) on achievement. Lastly, a set of moderator variables (instructional and demographic) are examined for their potential relationship with S-C learning and achievement outcomes.

Summary of Main Results

The following summarized points will be discussed in the proceeding "Authors' Conclusions" section below.

Overall Test of Student-Centered Instruction.

Regarding the influence, and effectiveness, of more S-C learning vs. less S-C (more T-C) learning, as well as the degree of S-C learning, on undergraduate level science post-secondary achievement outcomes, two tests were used:

- The first examined the overall outcome of 141 effect sizes drawn from 96 sources. The results were significant, producing an average random effect of $\overline{g}$ = 0.34. This result would be considered a moderate effect favoring more S-C learning according to Cohen's (1988) interpretative criteria.
- The second examined the overall strength of the relationship between treatment and control (i.e., degree of student-centeredness). The results of the meta-regression

analysis were non-significant, producing a weak-to-moderate positive relationship (β = 0.28) between degree of S-C and degree of student achievement.

While the first test suggests that, on average, *more* undergraduate level S-C classroom environments produce better achievement outcomes than do *less* S-C classroom ones, the non-significant meta-regression result (p = 0.40) – even though a somewhat linear relationship – takes away from the overall strength of the claim that *increasing* the degree of S-C in classroom environments will subsequently *increase* the degree of student achievement relative to those classrooms engaging in less S-C (more T-C) instructional practices.

Primary Predictor Variables.

Regarding the predictive power of the four instructional dimensions (*Pacing, Teacher's Role, Flexibility,* and *Adaptation*) on achievement outcomes:

- Only *Flexibility* (via simple-regression) produced a significant negative relationship (p = 0.05).

Regarding the degree of S-C of each of the four dimensions and their individual magnitude of effect on student achievement outcomes:

- Only *Flexibility* (via mixed moderator variable analysis) produced a significant result (p = 0.05).
- *Flexibility* produced positive effect sizes at differential scores 0, +1 and +2, but produced negative effect sizes at differential scores +3 and +4.

The results from the mixed moderator variable analysis indicate no significant relationship between the other three instructional dimensions – *Pacing, Teacher's Role,* and *Adaptation* – and student achievement outcomes.

Instructional and Demographic Moderator Variables.

Three moderator variables – technology use, subject matter, and treatment group class size – were added to the analysis and were first analyzed via categorical-level analysis:

- Technology use:
    - Instructional settings where both participation groups did not use technology ($\bar{g}$ = 0.40), and settings where only the treatment group used technology ($\bar{g}$ = 0.48), vastly outperformed instructional settings where both groups used it ($\bar{g}$ = 0.19).
    - Across degrees of *Flexibility*, the instructional setting of both groups not using technology produced the only significant model (p = 0.03). Within this model (i.e., across degrees of *Flexibility* for both participation groups not using technology) a negative relationship was present – as *Flexibility* increased, student achievement outcomes decreased.
- Subject matter:
    - Only chemistry ($\bar{g}$ = 0.58), physics ($\bar{g}$ = 0.45) and biology ($\bar{g}$ = 0.31) were significant.
- Class size:
    - Medium classes (16 to 49 students) and large classes (50 to 99 students) were significant and produced the highest effect sizes ($\bar{g}$ = 0.40 and $\bar{g}$ = 0.46, respectively).
    - Very large classes (100+ students) had the smallest effect ($\bar{g}$ = 0.09), which was non-significant.

The moderator variables were then analyzed via mixed moderator variable analysis:

- Chemistry was the best predictor of effect size ($\beta = 0.23$, $SE = 0.11$, $z = 2.13$, $p = 0.03$), with a significantly higher average effect size than the reference group of biology.
- Studies in which both groups used technology had a significantly lower average effect size ($\beta = -0.18$, $SE = 0.09$, $z = -2.09$, $p = 0.04$) than the reference group of studies in which both groups did not use technology.
- Studies in which the treatment group used technology and the control group did not, the result was not significantly different from studies in which both groups did not use technology ($\beta = 0.07$, $SE = 0.11$, $z = 0.06$, $p = 0.53$). In other words, the presence of technology in the treatment group does not outperform studies where no technology is in either group.

The summarized points from the above section will be discussed in the "Authors' Conclusions" section below.

Overall Completeness and Quality of the Evidence

While this systematic review did explore instructional events occurring within undergraduate level science classrooms, it was not able to consider all instances of literature covering between-group undergraduate level science classroom comparisons. Such a goal would have been extremely ambitious in nature. As a result, the search and retrieval process was selective in two important ways. First, we selected for sources that compared two groups and that contained enough individual group description to determine if the S-C qualities for which we were searching were present. Second, we only selected for high-quality QEDs and RCTs. Based on these two selection criteria, it was deemed that the 96 sources and 141 independent effect sizes are an appropriate representation of the larger body of sources that were either excluded or not able to be assessed.

Limitations and Potential Biases in the Review Process

The main limitation and potential bias from the overall review process is the use of *high-inference coding*. As outlined in the Methods section, the designation of treatment and control in each individual study is not based on the specific designations already provided by the authors, but rather two reviewers working independently basing it on a set of judgments on each of the four instructional dimensions (*Pacing, Teacher's Role, Flexibility,* and *Adaptation*). The condition rated higher in S-C (i.e., a higher total score on the sum of each dimension's individual score) is deemed the treatment, and the condition rated lower the control. It is important to note in considering the accuracy of coding, and as explained in the Methods section, that the reviewers received extensive training for this task, which included multiple practice runs on studies previously judged to have been accurately and reliably coded. The same two reviewers also had extensive experience from working on the coding for the Bernard et al. (2019) systematic review. The inter-rater agreement rate for the instructional dimensions coding in this systematic review was judged to be high with Cohen's $\kappa = 0.89$.

Another important note is that the overall research team under which this meta-analysis falls, and whose results will contribute to that research team's body of student-centered literature, has extensive experience with this particular process of establishing the treatment and control conditions from individual studies via high-inference coding. In particular, Borokhovski, Bernard, Tamim and Abrami (2009) presented a paper on the subject at the Campbell Collaboration's Ninth Colloquium.

While this form of *high-inference coding* does result in greater risk of bias than the standard treatment/control designations (i.e., *low-inference coding*), for the purposes of this systematic review it is deemed to be the only way to advance the research literature beyond relatively simple comparisons between 'either this or that' like the standard treatment/control designations that populate the educational research literature. As has been repeatedly mentioned before, this systematic review is concerned with the questions of "which," "when," and "for what purpose" regarding combinations of T-C and S-C learning approaches – thus reflecting a "greater-to-lesser" S-C learning scale along a continuum of instructional practices, rather than an "either/or" view of T-C and S-C learning.

Author's Conclusions

This systematic review provides evidence that student-centered instruction leads to increases in undergraduate level science student achievement outcomes, and thus to greater increases in learning. This is seen by the medium sized overall random effects average of $\overline{g}$ = 0.34. However, the argument is hindered by the fact that a significant, linear relationship between degree of S-C and degree of achievement was not found (β = 0.01, *SE* = 0.01, z = 0.03, p = 0.40). As a result, the simple argument that increasing the amount of S-C learning in an undergraduate level science course will in turn increase student achievement cannot so easily be made. The relationship between S-C learning and student achievement in this particular instructional setting is more nuanced, and factors such as type of and degree of S-C learning, as well as combinations of S-C learning and combinations of S-C and T-C learning, are not without their roles and impact. For instance, and as will be discussed later in this section, variables such as class size, specific science subject matters (and their specific course syllabi) within undergraduate level science, as well as the type of, and amount of, technology used in conjunction with the degree of the flexibility of classroom instruction, all can be shown to influence, in different directions, the relationship between S-C learning and student achievement outcomes.

The above two results from the overall test of S-C instruction, as well as variables such as the ones mentioned above, also further lend to the importance of the instructional setting in determining the specific type(s) of, and impact of, more S-C learning. For example, in the systematic review by Bernard et al. (2019) on S-C instruction in K-12, not only was the overall effect in favour of S-C learning ($\bar{g}$ = 0.44), but also a significant (p = 0.03) positive linear relationship between degree of S-C and degree of achievement outcomes was present. In that particular instructional setting, increasing the amount of S-C learning was more conducive to improvements in achievement compared to the improvements in undergraduate level science instructional settings.

When considering the instructional settings in which undergraduate science courses take place, explanations for the results of the primary predictor variable analyses can be inferred. The fact that *Pacing*, *Teacher's Role* and *Adaptability* were all non-significant in both meta-regression and mixed moderator variable analysis could be indicative of the typical classroom structure of undergraduate chemistry, biology, physics, geology, and psychology courses being more T-C based. Courses of this nature tend to be larger, particularly in the more introductory levels, and they also tend to be predominantly lecture-based. The content in these courses also does not drastically change from year to year, as many of the courses are centered on longstanding established theories, principles, and historical findings. As a result, course syllabi do not experience major changes year to year in terms of the content covered. Furthermore, the syllabus received by students at the start of a semester will seldom see changes, save for perhaps the changing of the presentation dates of certain topics, or the rescheduling of a lecture/topic due to, for example, a lecture being cancelled because of unforeseen circumstances. Many course syllabi are also in conjunction with the official textbook being used for the course, and many professors are even provided with official lecture slides by the company who manufactured the course textbook. In short, the typical undergraduate level science course might not always afford significant opportunities for the insertion of more S-C based instructional dimensions such as *Pacing*, *Teacher's Role* and *Adaptability*, nor might the learning environment be particularly conducive to the implementation and success of these dimensions in terms of improving student learning achievement outcomes. It is important to note that the majority of the studies in this systematic review did not score higher than +2 on the four instructional dimensions, with most scoring between 0 to +2. This raises the potential limitation of dealing with a particular sample – in this case undergraduate level science – that might be low in S-C to begin with, and as such the instances, and potential subsequent positive effects, of high levels of S-C learning (+3 and +4) in turn might be negligible.

While the operationalization of *Flexibility* in the Methods section also has aspects that would enable it to be rated more on the T-C side of the spectrum in undergraduate level science settings, there are certain aspects of the science subject matters chosen for this meta-analysis that could help explain why, of the four dimensions, it was the only significant (albeit negative) predictor of effect size. *Flexibility* in part reflects the "degree of student control over progression through the course content" and the laboratory component of science courses such as chemistry, biology, and physics could provide the particular instructional setting that reflects an increased degree of student control in this aspect (i.e., increased S-C learning). Labs are often a mandatory component of undergraduate level science courses, especially in more introductory courses, and within labs students have the ability to move at their own pace as they work through the assigned work. Of course, students are still working within the overall parameters of the length of time allotted to finish and hand in work, and labs are still within the overall, pre-established course syllabus, however different students (or students and their partners if it is paired work) can finish the same work in different times, as well as approach the work in different manners depending on their individual preferences and strategies.

Of the five subject matters that were coded for, chemistry ($\bar{g}$ = 0.58), physics ($\bar{g}$ = 0.45) and biology ($\bar{g}$ = 0.31) all had much higher effect sizes than psychology ($\bar{g}$ = 0.06) and geology ($\bar{g}$ = -0.05). Unlike psychology and geology, the effect sizes of these three subject matters were also significant. Of the five subject matters, chemistry, physics and biology tend to have the mandatory lab components, which could very well be a key contributing factor to their significant, positive effect sizes. An important point to consider however is whether it is the actual S-C aspects of labs that are influencing the dependent variable of student achievement outcomes, or does the influence stem from the pedagogical nature itself of the use of labs (and this very pedagogical nature is actually being drowned out by the general, positive influence of S-C learning)? It was not possible to separate out the presence of labs from the overall curriculum vis-à-vis S-C, so questions remain as to their role in science courses. This is discussed below.

While a laboratory component could be conducive to increased *Flexibility*, other aspects of a science course setting (i.e., class size, lecture-based, rigid syllabus) could result in that same instructional dimension exemplifying T-C components. As mentioned earlier, the significant relationship between degree of *Flexibility* and effect size was negative, thus indicating that only up to a certain point will increasing *Flexibility* benefit student achievement outcomes. Implementing too much *Flexibility* – like with too much *Pacing*, *Teacher's Role* and *Adaptability* – could also render this instructional setting less conducive to increases in student achievement outcomes. To once again echo the same point: improving student achievement outcomes is more complicated than simply increasing S-C learning.

Very interesting findings concerned the moderator variable of technology use. Studies in which both participation groups used technology ($\bar{g}$ = 0.19) were significantly outperformed by studies in which either both groups did not use technology ($\bar{g}$ = 0.40) or the treatment group did and the control group did not ($\bar{g}$ = 0.48). Additionally, meta-regression analysis showed no significant difference between both groups not using technology and only the treatment group using it – the presence of technology only in the treatment group did not outperform studies where no technology was present in either group. Lastly, the degree of technology use across degrees of *Flexibility* was only significant (p = 0.03) when both groups did not use technology, and there was a negative relationship – more *Flexibility* resulted in lower achievement outcomes when no technology was present in both participation groups.

The fact that settings in which only the treatment group used technology were not significantly different than settings in which both groups did not use technology could be due to the type of technology being used. The technology found in typical undergraduate level science instructional settings includes PowerPoint slides with the occasional video during lectures, as well as some form of an online course management system such as Moodle and Blackboard. A meta-analysis conducted by Schmid et al. (2014) on the effects of technology use on achievement in post-secondary education looked into the different pedagogical uses of technology, and found that technological applications that simply present information in an alternative form, such as PowerPoint, yielded small effects in the range of $0.10 < \bar{g} < 0.20$ (i.e., what would be considered 'trivial' in social sciences). The fact that this particular application of technology does not incur much of an advantage in terms of achievement outcomes could be an explanation as to why no significant differences between instructional settings with no technology and settings with only treatment using it exist. However, the above does not explain why achievement outcomes in instructional settings in which both groups used technology is significantly lower than when at least one group (i.e., at least the control group) did not use it. If this particular form of technology use does incur benefits, albeit small, as reported by Schmid et al. (2014), then why does adding technology to the control group significantly diminish effect size so much?

Another angle from which to approach the relationship between technology use and effect size is by considering how the technology supports the selected pedagogy. Does the pedagogical use of technology mentioned above - alternative forms of presenting information – more so support S-C learning, or T-C learning? Is this interaction further complicated by the nature of the content and learners' prior knowledge? Again, see the future research section below.

The finding of undergraduate level science instructional settings with both groups not using technology significantly outperforming settings in which both groups used it goes against the finding of Schmid et al. (2014) that settings where the control groups used some technology ($\bar{g}$ = 0.31) performed better than settings where the control groups used no technology ($\bar{g}$ = 0.25). However, the Schmid et al. (2014) meta-analysis covered a larger range of subject matters, including all forms of STEM as well as non-STEM subjects. What does match findings by

Schmid et al. (2014) is the fact that STEM settings in which the control group did not use technology resulted in a positive relationship with achievement, whereas STEM settings in which the control group used only a bit of technology resulted in a negative relationship. This goes with the finding that undergraduate level science settings which contain no technology in the control group outperformed science settings which did contain technology in the control group. However, yet again, the subject matter covered in Schmid et al. (2014) was broader, encompassing all of STEM instead of just science.

In summary, while the overall model suggests that S-C learning does result in student learning achievement outcome improvements ($\overline{g}$ = 0.34), the effects are most pronounced when considering the instructional dimension of *Flexibility*, as well as when considering the potential impact of laboratory components in science courses such as biology, physics, and in particular chemistry. It is important to remember that the discussion about labs is merely speculation – the specific course content, and syllabi, of the science courses in the studies used for this systematic review were not examined. It is also important to remember that a consideration of how the pedagogy itself in these instructional settings influences student achievement outcomes, as well as how factors such as technology support the selected pedagogy, is warranted in the quest to better understand the relationship between instructional settings and student learning achievement outcomes.

Implications for Practice and Policy

While this meta-analysis does not provide an instructional design blueprint for creating more S-C undergraduate level science classrooms, it does point to certain areas where types of, and degree of, S-C learning can be applied for the benefit of increasing student learning achievement outcomes. As previously mentioned, instructional settings favouring the dimension of *Flexibility* (up to a certain point), and courses with lab components, particularly chemistry, could potentially be the most conducive to achievement outcome benefits stemming from S-C learning practices. As a whole, developing the optimal environments for undergraduate level science learning to take place needs to consider the specifics and nuances that go into the creation of these environments, rather than simply taking a surface-level, generalized approach and analysis of information.

Implications for Future Research

In line with previous research on examining instructional events in isolation in an attempt to veer away from an "either/or" view of T-C and S-C learning, and towards a "greater-to-lesser" S-C learning scale along a continuum of instructional practices (Bernard et al. 2013; Bernard et al. 2019), this systematic review also seeks to argue that more nuanced questions in classroom-based research need to become the norm. Different types of instructional settings present different challenges and different opportunities for improving the learning experience, and as such future research endeavours, and future research questions, need to be tailored to these differences as much as possible.

Additionally, more tailored future systematic reviews will benefit from primary researchers reporting as much of the information from their studies as possible, and in sufficient enough detail, for the purpose of their studies in turn being used to advance the research literature. In conducting this systematic review, many promising studies were excluded due to there not being enough information as to what the treatment and control groups did during their respective classroom interventions. Sufficient detail in reporting can also help future systematic reviews explore potential explanations for findings.

Regarding future implications stemming from this particular meta-analysis, a better understanding of the impact of technology use on student achievement outcomes is a worthwhile undertaking. A better understanding of not only the different types of technology use, but also the degree of technology use as well as the degree of difference in amount of technology use between participation groups can further move the narrative of developing optimal undergraduate level science instructional settings forward. A better understanding also largely involves further delving into technology's functionality in these particular instructional settings. In short, how does the selected technology support the selected pedagogy?

Regarding the four instructional dimensions explored - *Pacing, Teacher's Role, Flexibility, and Adaptation* – additional research needs to consider the potential impact of other instructional dimensions and their combinations. Bernard et al. (2013) performed a meta-analysis in K-12 with a total of 11 instructional dimensions, and research in undergraduate level science settings can benefit from an exploration of how (or even if) these additional dimensions impact achievement.

A final avenue for future research concerns a more in-depth look at the role of laboratories in undergraduate level science classrooms and the potential types of, and degree of, S-C learning taking place within this particular sub-setting of a larger instructional setting (i.e., labs being a part of a science course as a whole). The role of labs in science courses, and their potential relationship with student achievement outcomes, should also explored from a pedagogical perspective to better determine if the dependent variable of achievement outcomes is more a function of S-C instruction or pedagogy. Additionally, the instructional dimension of *Flexibility* and its relationship with labs deserves an in-depth exploration in the quest for the development of optimal undergraduate level science post-secondary education learning environments.

References

Aiello, N. C., & Wolfle, L. M. (1980). *A meta-analysis of individualized instruction in science*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. Retrieved from http://files.eric.ed.gov/fulltext/ED190404.pdf

Aiken, W. M. (1942). *The story of the eight-year study. Adventures in American education* (Vol. 1). New York: McGraw-Hill.

Bernard, R. M., Borokhovski, E., et al. (2013, April). *Teacher-centered and Student-centered pedagogy: A meta-analysis of classroom practices and processes.* Poster presented at the 2013 meeting of the American Educational Research Association, San Francisco, CA.

Bernard, R. M., Borokhovski, E., Schmid, R. F., Waddington, D. I., & Pickup, D. (2019). Twenty-first century adaptive teaching and individualized learning operationalized as specific blends of student-centered instructional events: A systematic review and meta-analysis, *Campbell Systematic Reviews*. 15:e1017. https://doi.org/10.1002/c12.1017

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive Meta-analysis Version 3.* Englewood, NJ: Biostat.

Borokhovski, E., Bernard, R. M., Tamim, R., & Abrami, P. C. (2009, May). *Establishing the direction of effect in meta-analyses with multiple treatments (and no obvious control condition).* Paper presented at the Campbell Collaboration Ninth Annual Colloquium, Oslo, Norway, May 18-20, 2009.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dewey, J. (1938). *Experience and education.* New York: Simon & Schuster.

Duvall, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Gersten, R., Chard, D., Jayanthi, M., Baker, S., Morphy, P., & Flojo, J. (2008). *Mathematics instruction for students with learning disabilities or difficulty learning mathematics: A synthesis of the intervention research.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Jonassen, D. H. (1991). Evaluating constructivist learning. *Educational Technology, 31*(9), 28–33.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 9–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Morrison, H. C. (1926). *The practice of teaching in the secondary school.* Chicago, IL: University of Chicago Press.

National Mathematics Advisory Panel (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education.

Rosenshine, B. (1976). Classroom instruction. In N. L. Gage (Ed.), *The psychology of teaching methods* (75th NSSE Yearbook) (pp. 335–371). Chicago, IL: University of Chicago Press.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 376–391). New York: Macmillan.

Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A.,…Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education, 72,* 271-291.

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In Tobias, S., & Duffy, T. M. (Eds.), *Constructivist instruction: Success or failure* (pp. 34-61). Routledge: New York, NY.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71.

Tobias, S., & Duffy, T. M. (Eds.) (2009). Constructivist instruction: Success or failure. Routledge: New York, NY.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Pres

Appendix 1
Descriptive Statistics for Each Study

Study Name	Date	Hedges' *g*	*SE*-g	Confidence Interval and Test Statistics			
				Lower *CI*	Upper *CI*	*z*-Value	*p*-Value
Alcazar	2005	0.081	0.080	-0.076	0.239	1.010	0.312
Altiparmak	2009	0.893	0.318	0.270	1.517	2.808	0.005
Ambrosio	1993	0.112	0.326	-0.526	0.750	0.344	0.731
Andrews	1984	1.248	0.290	0.680	1.816	4.306	0.000
Arburn	1999	0.085	0.244	-0.392	0.563	0.350	0.726
Armstrong	2007	0.053	0.058	-0.060	0.167	0.922	0.357
Atan	2005	0.358	0.261	-0.154	0.869	1.370	0.171
Azevedo	2004	0.514	0.176	0.170	0.859	2.925	0.003
Azevedo	2004	1.013	0.358	0.312	1.715	2.830	0.005
Barab	2009	1.383	0.439	0.523	2.242	3.152	0.002
Barak	2005	0.696	0.138	0.425	0.968	5.029	0.000
Basili	1991	0.506	0.258	0.000	1.012	1.959	0.050
Beam	2010	-0.114	0.437	-0.971	0.743	-0.261	0.794
Bechtel	1963	-0.258	0.251	-0.750	0.233	-1.032	0.302
Bilgin	2009	0.406	0.228	-0.040	0.852	1.786	0.074
Bilgin	2006	1.050	0.218	0.622	1.478	4.812	0.000
Cacciatore	2009	0.377	0.250	-0.114	0.867	1.504	0.133
Cahyadi_1	2004	1.076	0.184	0.715	1.437	5.839	0.000
Cahyadi_1	2007	0.401	0.205	0.000	0.803	1.958	0.050
Cahyadi_2	2004	0.833	0.166	0.508	1.158	5.027	0.000
Cahyadi_2	2007	0.636	0.206	0.231	1.040	3.082	0.002
Caldwell	1978	0.092	0.136	-0.175	0.360	0.676	0.499
Cayton_1	1975	0.036	0.324	-0.599	0.672	0.112	0.911
Cayton_2	1975	-0.012	0.318	-0.635	0.612	-0.037	0.971
Chou	1998	0.435	0.269	-0.092	0.962	1.618	0.106
Corbalan	2009	0.042	0.260	-0.469	0.552	0.160	0.873
Cox	2002	0.383	0.251	-0.109	0.876	1.527	0.127
Davies_1	1981	0.603	0.212	0.187	1.018	2.845	0.004
Davies_2	1981	0.849	0.290	0.280	1.418	2.924	0.003
Demastes	1995	0.007	0.146	-0.278	0.293	0.051	0.959
Dori_1	2005	0.723	0.283	0.168	1.278	2.555	0.011
Dori_2	2005	0.405	0.161	0.091	0.720	2.524	0.012
Dori_3	2005	0.368	0.150	0.074	0.663	2.449	0.014
Doymus	2010	-0.449	0.228	-0.897	-0.002	-1.968	0.049
Doymus_1	2008	1.000	0.349	0.315	1.685	2.863	0.004
Doymus_2	2008	1.725	0.265	1.206	2.244	6.514	0.000
Elshout_1	1992	0.838	0.751	-0.634	2.310	1.116	0.265
Elshout_2	1992	-0.015	0.632	-1.255	1.224	-0.024	0.981

Elshout_3	1992	0.063	0.606	-1.124	1.251	0.105	0.917
Elshout_4	1992	-0.340	0.714	-1.741	1.060	-0.477	0.634
Emerson_1	1988	0.572	0.292	-0.001	1.144	1.957	0.050
Emerson_2	1988	-0.193	0.289	-0.760	0.374	-0.668	0.504
Evans	2008	0.421	0.301	-0.168	1.011	1.400	0.162
Ezrailson	2004	0.281	0.265	-0.238	0.800	1.062	0.288
Falconer_1	2001	1.297	0.208	0.888	1.705	6.222	0.000
Falconer_2	2001	0.991	0.173	0.653	1.330	5.746	0.000
Falconer_3	2001	1.452	0.318	0.829	2.074	4.568	0.000
Feldo	2010	0.073	0.126	-0.175	0.320	0.575	0.565
Franklin	1994	0.260	0.137	-0.009	0.530	1.894	0.058
Friedel	2008	0.337	0.290	-0.231	0.905	1.163	0.245
Gifford	1982	0.628	0.226	0.185	1.071	2.778	0.005
Gossman	2007	-0.227	0.361	-0.934	0.480	-0.630	0.529
Hall	1990	0.257	0.184	-0.102	0.617	1.402	0.161
Hill	1999	-0.172	0.448	-1.050	0.706	-0.383	0.701
Hulshof	2005	-0.717	0.241	-1.190	-0.245	-2.977	0.003
Ibrahim	2001	1.214	0.235	0.753	1.674	5.169	0.000
Jones	1980	-0.271	0.182	-0.629	0.086	-1.488	0.137
Kapp	2011	-0.689	0.141	-0.964	-0.413	-4.896	0.000
Knight	2005	0.532	0.169	0.201	0.863	3.153	0.002
Koenig	2007	0.586	0.172	0.249	0.923	3.404	0.001
Kremer	1991	-0.202	0.302	-0.794	0.391	-0.667	0.505
Lee_1	2010	-0.107	0.156	-0.412	0.199	-0.683	0.494
Lee_2	2010	0.654	0.156	0.348	0.961	4.187	0.000
Lee_3	2010	-0.262	0.156	-0.568	0.043	-1.683	0.092
Lee_4	2010	0.069	0.155	-0.235	0.374	0.448	0.654
LeTexier	2009	0.810	0.244	0.332	1.288	3.322	0.001
Levinson_1	2007	0.296	0.263	-0.220	0.812	1.123	0.261
Levinson_2	2007	-0.431	0.264	-0.948	0.087	-1.630	0.103
Lewis	2005	0.381	0.145	0.097	0.665	2.629	0.009
Liang_1	2005	0.138	0.310	-0.469	0.745	0.445	0.656
Liang_2	2005	-0.012	0.323	-0.644	0.620	-0.038	0.970
Liang_3	2005	-0.341	0.333	-0.994	0.313	-1.023	0.306
Lord	2006	0.344	0.205	-0.057	0.746	1.682	0.093
Martin	2007	0.571	0.196	0.187	0.954	2.914	0.004
Martin_1	2009	-1.039	0.453	-1.927	-0.150	-2.291	0.022
Martin_2	2009	-1.006	0.490	-1.966	-0.046	-2.054	0.040
Mathew	2008	1.308	0.329	0.662	1.953	3.971	0.000
McKee_1	2007	0.609	0.399	-0.173	1.392	1.526	0.127
McKee_2	2007	0.274	0.410	-0.529	1.077	0.669	0.503
McKee_3	2007	0.209	0.448	-0.670	1.087	0.466	0.641
McLaren	2009	0.214	0.194	-0.166	0.594	1.104	0.270
Moreno	2009	-0.238	0.266	-0.759	0.283	-0.897	0.370

Moreno_1	2004	1.002	0.297	0.421	1.584	3.381	0.001
Moreno_2	2004	0.740	0.275	0.201	1.278	2.692	0.007
Morgil_1	2006	1.080	0.225	0.639	1.521	4.797	0.000
Morgil_2	2006	0.718	0.221	0.284	1.152	3.245	0.001
Morris	1978	0.006	0.177	-0.342	0.353	0.031	0.975
Muller_1	2008	0.443	0.172	0.106	0.780	2.578	0.010
Muller_2	2008	0.647	0.244	0.169	1.125	2.650	0.008
Munyofu_1	2008	-0.137	0.256	-0.639	0.366	-0.533	0.594
Munyofu_2	2008	0.330	0.259	-0.178	0.838	1.274	0.203
Munyofu_3	2008	0.360	0.259	-0.148	0.868	1.388	0.165
Munyofu_4	2008	0.455	0.260	-0.055	0.964	1.750	0.080
Munyofu_5	2008	0.423	0.260	-0.086	0.932	1.629	0.103
Munyofu_6	2008	0.372	0.262	-0.141	0.884	1.421	0.155
Nokes	2005	0.295	0.257	-0.209	0.798	1.148	0.251
Nugent	2008	0.754	0.420	-0.069	1.578	1.795	0.073
Nugent	2012	0.404	0.260	-0.106	0.914	1.553	0.120
Okebukola_1	1988	1.512	0.560	0.414	2.610	2.700	0.007
Okebukola_2	1988	1.985	0.806	0.405	3.564	2.463	0.014
Okebukola_3	1988	0.066	0.570	-1.052	1.183	0.115	0.908
Okebukola_4	1988	1.661	0.481	0.719	2.603	3.456	0.001
Olajide_1	2010	0.906	0.294	0.329	1.482	3.079	0.002
Olajide_2	2010	0.846	0.303	0.253	1.440	2.794	0.005
Olson_1	1962	-0.252	0.298	-0.836	0.332	-0.846	0.398
Olson_2	1962	-0.128	0.292	-0.700	0.444	-0.439	0.661
Perry	2008	-0.392	0.289	-0.958	0.174	-1.356	0.175
Phelps	2012	0.253	0.152	-0.045	0.551	1.663	0.096
Proske	2012	0.247	0.313	-0.365	0.860	0.791	0.429
Quitadamo	2008	0.390	0.134	0.127	0.653	2.908	0.004
Quitadamo	2007	0.343	0.114	0.120	0.566	3.010	0.003
Reinhardt	2012	-0.852	0.290	-1.421	-0.283	-2.937	0.003
Ruiter	1971	0.718	0.252	0.224	1.212	2.849	0.004
Selcuk	2010	1.209	0.430	0.367	2.051	2.814	0.005
Senocak	2007	0.181	0.199	-0.209	0.571	0.909	0.363
Slish	2005	0.549	0.280	-0.001	1.099	1.958	0.050
Spivey_1	1995	0.083	0.316	-0.537	0.703	0.261	0.794
Spivey_2	1995	1.398	0.341	0.730	2.065	4.103	0.000
Stark	2009	0.586	0.288	0.021	1.152	2.033	0.042
Stiller_1	2009	0.406	0.274	-0.132	0.944	1.479	0.139
Stiller_2	2009	-0.119	0.267	-0.644	0.405	-0.447	0.655
Stout	1978	0.314	0.265	-0.205	0.833	1.185	0.236
Strawitz	1987	0.401	0.406	-0.395	1.197	0.988	0.323
Struyven	2010	-0.174	0.112	-0.393	0.045	-1.556	0.120
Sturges	2009	0.679	0.132	0.421	0.937	5.154	0.000
Suits	2004	0.554	0.193	0.176	0.931	2.875	0.004

Swanson_1	1990	-0.283	0.616	-1.490	0.924	-0.460	0.646
Swanson_2	1990	-0.385	0.619	-1.598	0.827	-0.623	0.533
Swanson_3	1990	-0.668	0.631	-1.904	0.569	-1.058	0.290
Tarhan	2012	1.393	0.350	0.706	2.079	3.974	0.000
van den Boom_1	2007	-0.127	0.417	-0.944	0.690	-0.304	0.761
van den Boom_2	2007	0.728	0.433	-0.120	1.576	1.683	0.092
Veenman_1	1994	-0.234	0.283	-0.790	0.321	-0.827	0.408
Veenman_2	1994	0.070	0.286	-0.490	0.630	0.244	0.808
Vreven	2007	-0.137	0.114	-0.360	0.085	-1.208	0.227
Walker	2008	0.242	0.089	0.067	0.417	2.710	0.007
Wittwer	2010	-0.484	0.369	-1.207	0.240	-1.311	0.190
Wozniak	2012	2.286	0.997	0.331	4.240	2.292	0.022
Wright	2008	-0.227	0.149	-0.519	0.065	-1.526	0.127
Wright	2006	-0.227	0.149	-0.519	0.065	-1.526	0.127
Yeo	2002	0.368	0.202	-0.028	0.765	1.820	0.069

Appendix 2
References to Included Sources

*Alcazar, M. T. M., & Fitzgerald, V. L. (2005). An experimental design to study the effectiveness of pbl in higher education, in first year science students at a university in Peru, South America. *College Quarterly, 8*(2). Retrieved from http://collegequarterly.ca/2005-vol08-num02-spring/alcazar_fitzgerald.html

*Altiparmak, M., & Nakiboglu Tezer, M. (2009). Hands on group work paper model for teaching DNA structure, central dogma and recombinant DNA. *US-China Education Review, 6*(1), 19-23.

*Andrews, J. D. W. (1984). Discovery and expository learning compared: Their effects on independent and dependent students. *Journal of Educational Research, 78*(2), 80-89. http://dx.doi.org/10.1080/00220671.1984.10885578

*Arburn, T. M., & Bethel, L. J. (1999, March). *Assisting at-risk community college students: Acquisition of critical thinking learning strategies*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.

*Armstrong, N., Chang, S.-M., & Brickman, M. (2007). Cooperative learning in industrial-sized biology classes. *CBE - Life Sciences Education, 6*(2), 163-171. http://dx.doi.org/10.1187/cbe.06-11-0200

*Atan, H., Sulaiman, F., & Idrus, R. M. (2005). The effectiveness of problem-based learning in the web-based environment for the delivery of an undergraduate physics course. *International Education Journal, 6*(3), 430-437.

*Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523-535. http://dx.doi.org/10.1037/0022-0663.96.3.523

*Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29*(3), 344-370. http://dx.doi.org/10.1016/j.cedpsych.2003.09.002

*Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology, 18*(4), 305-320. http://dx.doi.org/10.1007/s10956-009-9171-5

*Barak, M., & Dori, Y. J. (2005). Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment. *Science Education, 89*(1), 117-139. http://dx.doi.org/10.1002/sce.20027

*Basili, P. A., & Sanford, J. P. (1991). Conceptual change strategies and cooperative group work in chemistry. *Journal of Research in Science Teaching, 28*(4), 293-304. http://dx.doi.org/10.1002/tea.3660280403

*Beam, M. (2010). *A comparison of didactic and inquiry teaching methods in a rural community college earth science course* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3440772)

*Bechtel, L. P. (1963). *Comparative effects of differentiated teaching methods on certain personality characteristics of college students: The effect of the traditional approach to teaching psychology as compared to an interpersonal approach to teaching psychology upon beliefs, attitudes, values and adjustment of college students in a course in general psychology* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 302095386)

*Bilgin, I. (2006). The effects of pair problem solving technique incorporating Polya's problem solving strategy on undergraduate students' performance in chemistry. *Journal of Science Education, 7*(2), 101-106.

*Bilgin, I., Senocak, E., & Sozbilir, M. (2009). The effects of problem-based learning instruction on university students' performance of conceptual and quantitative problems in gas concepts. *EURASIA Journal of Mathematics, Science & Technology Education, 5*(2), 153-164.

*Burron, B., James, M. L., & Ambrosio, A. L. (1993). The effects of cooperative learning in a physical science course for elementary/middle level preservice teachers. *Journal of Research in Science Teaching, 30*(7), 697-707. http://dx.doi.org/10.1002/tea.3660300708

*Cacciatore, K. L., & Sevian, H. (2009). Incrementally approaching an inquiry lab curriculum: Can changing a single laboratory experiment improve student performance in general chemistry? *Journal of Chemical Education, 86*(4), 498-505.

*Cahyadi, V. (2004). The effect of interactive engagement teaching on student understanding of Introductory Physics at the Faculty of Engineering, University of Surabaya, Indonesia. *Higher Education Research and Development, 23*(4), 455-464. http://dx.doi.org/10.1080/0729436042000276468

*Cahyadi, V. (2007). Case 11: No more 'plug and chug'. In G. Crosling, L. Thomas & M. Heagney (Eds.), *Improving student retention in higher education: The role of teaching and learning* (pp. 131-138). London: Routledge. Retrieved from

*Caldwell, E. C. (1978). Mastery: The essential essential in PSI. *Teaching of Psychology, 5*(2), 59-65. http://dx.doi.org/10.1207/s15328023top0502_1

*Cayton, T. G. (1975). *A comparison of three methods of teaching child psychology* Available from ProQuest Dissertations and Theses database. (UMI No. 7506309)

*Chou, C.-H. (1998). *The effectiveness of using multimedia computer simulations coupled with social constructivist pedagogy in a college introductory physics classroom* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9839055)

*Corbalan, G., Kester, L., & J.G. van Merriënboer, J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction, 19*(6), 455-465. http://dx.doi.org/10.1016/j.learninstruc.2008.07.002

*Cox, A. J., & Junkin, W. F., III. (2002). Enhanced student learning in the introductory physics laboratory. *Physics Education, 37*(1), 37-44. http://dx.doi.org/10.1088/0031-9120/37/1/305

*Davies, C. S. (1981). Teaching introductory chemistry: Generality of the PSI approach. *Journal of Chemical Education, 58*(9), 686-689. http://dx.doi.org/10.1021/ed058p686

*Demastes, S. S., & et al. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching, 32*(5), 535-550. http://dx.doi.org/10.1002/tea.3660320509

*Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? *Journal of the Learning Sciences, 14*(2), 243-279.

*Doymus, K. (2008). Teaching chemical bonding through jigsaw cooperative learning. *Research in Science & Technological Education, 26*(1), 47-57. http://dx.doi.org/10.1080/02635140701847470

*Doymus, K., Karacop, A., & Simsek, U. (2010). Effects of jigsaw and animation techniques on students' understanding of concepts and subjects in electrochemistry. *Educational Technology Research and Development, 58*(6), 671-691. http://dx.doi.org/10.1007/s11423-010-9157-2

*Elshout, J. J., & Veenman, M. V. (1992). Relation between intellectual ability and working method as predictors of learning. *Journal of Educational Research, 85*(3), 134-143. http://dx.doi.org/10.1080/00220671.1992.9944429

*Emerson, I. (1988). A comparative evaluation of computer based and non-computer based instructional strategies. *Journal of Computers in Mathematics and Science Teaching, 8*(1), 46-49.

*Evans, K. L. (2008). *Learning stoichiometry: A comparison of text and multimedia instructional formats* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3284554)

*Ezrailson, C. M. (2004). *EMIT: Explicit modeling of interactive-engagement techniques for physics graduate teaching assistants and the impact on instruction and student performance in calculus-based physics* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3157053)

*Falconer, K., Wyckoff, S., Joshua, M., & Sawada, D. (2001, April). *Effect of reformed courses in physics and physical science on student conceptual understanding*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

*Feldon, D. F., Timmerman, B. C., Stowe, K. A., & Showman, R. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (cta) on lab report quality and student retention in the biological sciences. *Journal of Research in Science Teaching, 47*(10), 1165-1185. http://dx.doi.org/10.1002/tea.20382

*Franklin, G., Griffin, R., & Perry, N. (1994). Effects of cooperative tutoring on academic performance. *Journal of Educational Technology Systems, 23*(1), 13-25. http://dx.doi.org/10.2190/YEQT-UVF0-E1Q2-95RV

*Friedel, C., Irani, T., Rudd, R., Gallo, M., Eckhardt, E., & Ricketts, J. (2008). Overtly teaching critical thinking and inquiry-based learning: A comparison of two undergraduate biotechnology classes. *Journal of Agricultural Education, 49*(1), 72-84.

*Gifford, V. D., & Vicks, J. (1982). A comparison of the personalized system of instruction and a conventional biology course on the achievement of junior college freshmen. *Journal of Research in Science Teaching, 19*(8), 659-664. http://dx.doi.org/10.1002/tea.3660190805

*Gossman, P., Stewart, T., Jaspers, M., & Chapman, B. (2007). Integrating web-delivered problem-based learning scenarios to the curriculum. *Active Learning in Higher Education, 8*(2), 139-153. http://dx.doi.org/10.1177/1469787407077986

*Hall, D. A., & McCurdy, D. W. (1990). A comparison of a biological sciences curriculum study (bscs) laboratory and a traditional laboratory on student achievement at two private liberal arts colleges. *Journal of Research in Science Teaching, 27*(7), 625-636. http://dx.doi.org/10.1002/tea.3660270703

*Hill, L. A. (1999). *The implementation of learning strategies in the design of an on-line medical course supplement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9926696)

*Hulshof, C. D., Eysink, T. H. S., Loyens, S., & de Jong, T. (2005). ZAPs: Using interactive programs for learning psychology. *Interactive Learning Environments, 13*(1-2), 39-53. http://dx.doi.org/10.1080/10494820500224079

*Ibrahim, H. A. (2001). *Examining the impact of the Guided Constructivist teaching method on students' misconceptions about concepts of Newtonian physics* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3029041)

*Jones, N. W. (1980). A comparison between lecture and self-paced general geology. *Journal of Geological Education, 28*(5), 235-237. http://dx.doi.org/10.5408/0022-1368-28.5.235

*Kapp, J. L., Slater, T. F., Slater, S. J., Lyons, D. J., Manhart, K., Wehunt, M. D., & Richardson, R. M. (2011). Impact of redesigning a large-lecture introductory earth science course to increase student achievement and streamline faculty workload. *Journal of College Teaching and Learning, 8*(4), 23-35.

*Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education, 4*(4), 298-310. http://dx.doi.org/10.1187/05-06-0082

*Koenig, K. M., Endorf, R. J., & Braun, G. A. (2007). Effectiveness of different tutorial recitation teaching methods and its implications for ta training. *Physical Review Special Topics - Physics Education Research, 3*(1), 010104. http://dx.doi.org/10.1103/PhysRevSTPER.3.010104

*Kremer, J. F., & Dietzen, L. L. (1991). Two approaches to teaching accurate empathy to undergraduates: Teacher-intensive and self-directed. *Journal of College Student Development, 32*(1), 69-75.

*Lee, Y.-J. (2010). Effects of instructional preparation strategies on problem solving in a web-based learning environment. *Journal of Educational Computing Research, 42*(4), 385-406.

*LeTexier, K. (2009). *Storytelling as an active learning strategy in introduction to psychology courses* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3320287)

*Levinson, A. J., Weaver, B., Garside, S., McGinn, H., & Norman, G. R. (2007). Virtual reality and brain anatomy: A randomised trial of e-learning instructional designs. *Medical Education, 41*(5), 495-501. http://dx.doi.org/10.1111/j.1365-2929.2006.02694.x

*Lewis, S. E., & Lewis, J. E. (2005). Departing from lectures: An evaluation of a peer-led guided inquiry alternative. *Journal of Chemical Education, 82*(1), 135. http://dx.doi.org/10.1021/ed082p135

*Liang, L. L., & Gabel, D. L. (2005). Effectiveness of a constructivist approach to science instruction for prospective elementary teachers. *International Journal of Science Education, 27*(10), 1143-1162. http://dx.doi.org/10.1080/09500690500069442

*Lord, T., & Orkwiszewski, T. (2006). Moving from didactic to inquiry-based instruction in a science laboratory. *American Biology Teacher, 68*(6), 342-345. http://dx.doi.org/10.1662/0002-7685(2006)68[342:DTIIIA]2.0.CO;2

*Martin, M. (2009). *The effect of active techniques combined with didactic lecture on student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1467105)

*Martin, T., Rivale, S. D., & Diller, K. R. (2007). Comparison of student learning in challenge-based and traditional instruction in biomedical engineering. *Annals of Biomedical Engineering, 35*(8), 1312-1323. http://dx.doi.org/10.1007/s10439-007-9297-7

*Mathew, E. (2008). *Learning physics: A comparative analysis between instructional design methods* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3307901)

*McKee, E., Williamson, V. M., & Ruebush, L. E. (2007). Effects of a demonstration laboratory on student learning. *Journal of Science Education and Technology, 16*(5), 395-400. http://dx.doi.org/10.1007/s10956-007-9064-4

*McLaren, I. A. M., & Webber, D. (2009). Writing right: Enhancing student engagement and performance in an ecology course. *International Journal of Environmental and Science Education, 4*(4), 365-380.

*Moreno, R. (2009). Constructing knowledge with an agent-based instructional program: A comparison of cooperative and individual meaning making. *Learning and Instruction, 19*(5), 433-444. http://dx.doi.org/10.1016/j.learninstruc.2009.02.018

*Morgil, I., Penn, J. H., Secken, N., & Oskay, O. O. (2006). Introduction to complexation and masking within a computer-enriched module for analytical chemistry. *Turkish Online Journal of Distance Education, 7*(2). Retrieved from http://tojde.anadolu.edu.tr/makale_goster.php?id=261

*Morris, E. K., Surber, C. F., & Bijou, S. W. (1978). Self- versus instructor-pacing: Achievement, evaluations, and retention. *Journal of Educational Psychology, 70*(2), 224-230. http://dx.doi.org/10.1037/0022-0663.70.2.224

*Muller, D. A., Sharma, M. D., & Reimann, P. (2008). Raising cognitive load with linear multimedia to promote conceptual change. *Science Education, 92*(2), 278-296. http://dx.doi.org/10.1002/sce.20244

*Munyofu, M. (2008). *Effects of varied enhancement strategies (chunking, feedback, gaming) in complementing animated instruction in facilitating different types of learning objectives* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3414357)

*Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned? *Cognitive Science, 29*(5), 769-796. http://dx.doi.org/10.1207/s15516709cog0000_32

*Nugent, G., Kunz, G., Levy, R., Harwood, D., & Carlson, D. (2008). The impact of a field-based, inquiry-focused model of instruction on preservice teachers' science learning and attitudes. *Electronic Journal of Science Education, 12*(2), 94-111.

*Nugent, G., Toland, M. D., Levy, R., Kunz, G., Harwood, D., Green, D., & Kitts, K. (2012). The impact of an inquiry-based geoscience field course on pre-service teachers. *Journal of Science Teacher Education, 23*(5), 503-529. http://dx.doi.org/10.1007/s10972-012-9283-2

*Okebukola, P. A., & Jegede, O. J. (1988). Cognitive preference and learning mode as determinants of meaningful learning through concept mapping. *Science Education, 72*(4), 489-500. http://dx.doi.org/10.1002/sce.3730720408

*Olajide, J. O., & Adeoye, F. A. (2010). The effect of analogical reasoning and extended wait time on achievement in biology. *International Journal of Educational Sciences, 2*(1), 21-28. Retrieved from http://www.krepublishers.com/02-Journals/IJES/IJES-02-0-000-10-Web/IJES-02-1-000-10-Abst-PDF/IJES-02-1-21-10-027-Olajide-J-O/IJES-02-1-21-10-027-Olajide-J-O-Tt.pdf

*Olson, K. V. (1962). An experimental evaluation of a student-centered method and a teacher-centered method of biological science instruction for general education of college students. *Science Education, 46*(4), 367-373. http://dx.doi.org/10.1002/sce.3730460414

*Perry, J., Meir, E., Herron, J. C., Maruca, S., & Stal, D. (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE - Life Sciences Education, 7*(2), 193-201. http://dx.doi.org/10.1187/cbe.07-01-0007

*Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing. *Journal of Research in Reading, 35*(2), 136-152. http://dx.doi.org/10.1111/j.1467-9817.2010.01450.x

*Quitadamo, I. J., Faiola, C. L., Johnson, J. E., & Kurtz, M. J. (2008). Community-based inquiry improves critical thinking in general education biology. *CBE - Life Sciences Education, 7*(3), 327-337. http://dx.doi.org/10.1187/cbe.07-11-0097

*Quitadamo, I. J., & Kurtz, M. J. (2007). Learning to improve: Using writing to increase critical thinking performance in general education biology. *CBE - Life Sciences Education, 6*(2), 140-154. http://dx.doi.org/10.1187/cbe.06-11-0203

*Reinhardt, C. H., & Rosen, E. N. (2012). How much structuring is beneficial with regard to examination scores? A prospective study of three forms of active learning. *Advances in Physiology Education, 36*(3), 207-212. http://dx.doi.org/10.1152/advan.00108.2011

*Ruiter, W. W. (1971). *An analysis of the effectiveness of a learner-centered teaching system compared to that of a conventional teaching of basic electricity to university students* Available from ProQuest Dissertations and Theses database. (UMI No. 7125072)

*Selcuk, G. S. (2010). The effects of problem-based learning on pre-service teachers' achievement, approaches and attitudes towards learning physics. *International Journal of the Physical Sciences, 5*(6), 711-723.

*Senocak, E., Taskesenligil, Y., & Sozbilir, M. (2007). A study on teaching gases to prospective primary science teachers through problem-based learning. *Research in Science Education, 37*(3), 279-290. http://dx.doi.org/10.1007/s11165-006-9026-5

*Slish, D. F. (2005). Assessment of the use of the jigsaw method and active learning in non-majors, introductory biology. *Bioscene: Journal of College Biology Teaching, 31*(4), 4-10.

*Spivey, N. R. (1995). *Reciprocal teaching of lecture comprehension and comprehension monitoring skills in college students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9503727)

*Stark, R., Puhl, T., & Krause, U.-M. (2009). Improving scientific argumentation skills by a problem-based learning environment: Effects of an elaboration tool and relevance of student characteristics. *Evaluation & Research in Education, 22*(1), 51-68. http://dx.doi.org/10.1080/09500790903082362

*Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology, 25*(2), 184-203. http://dx.doi.org/10.14742/ajet.1149

*Stout, L. J. (1978). *A comparison of four different pacing strategies of Personalized System of Instruction and a traditional lecture format* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 7808600)

*Strawitz, B. M., & Malone, M. R. (1987). Preservice teachers' acquisition and retention of integrated science process skills: A comparison of teacher-directed and self-instructional strategies. *Journal of Research in Science Teaching, 24*(1), 53-60. http://dx.doi.org/10.1007/BF02357174

*Struyven, K., Dochy, F., & Janssens, S. (2010). "Teach as you preach": The effects of student-centred versus lecture-based teaching on student teachers' approaches to teaching. *European Journal of Teacher Education, 33*(1), 43-64. http://dx.doi.org/10.1080/02619760903457818

*Sturges, D., Maurer, T. W., & Cole, O. (2009). Understanding protein synthesis: A role-play approach in large undergraduate human anatomy and physiology classes. *Advances in Physiology Education, 33*(2), 103-110. http://dx.doi.org/10.1152/advan.00004.2009

*Suits, J. P. (2004). Assessing investigative skill development in inquiry-based and traditional college science laboratory courses. *School Science and Mathematics, 104*(6), 248-257. http://dx.doi.org/10.1111/j.1949-8594.2004.tb17996.x

*Tarhan, L., & Sesen, B. A. (2012). Jigsaw cooperative learning: Acid-base theories. *Chemistry Education Research and Practice, 13*(3), 307-313. http://dx.doi.org/10.1039/C2RP90004A

*van den Boom, G., Paas, F., & van Merrienboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction, 17*(5), 532-548. http://dx.doi.org/10.1016/j.learninstruc.2007.09.003

*Veenman, M. V. J., & Elshout, J. J. (1994). Differential effects of instructional support on learning in simulation environments. *Instructional Science, 22*(5), 363-383. http://dx.doi.org/10.1007/BF00891961

*Vreven, D., & McFadden, S. (2007). An empirical assessment of cooperative groups in large, time-compressed, introductory courses. *Innovative Higher Education, 32*(2), 85-92. http://dx.doi.org/10.1007/s10755-007-9040-1

*Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: Integrating active learning into a large lecture course. *CBE - Life Sciences Education, 7*(4), 361-367. http://dx.doi.org/10.1187/cbe.08-02-0004

*Walker, J. P., Sampson, V., Grooms, J., Anderson, B., & Zimmerman, C. O. (2012). Argument-driven inquiry in undergraduate chemistry labs: The impact on students' conceptual understanding, argument skills, and attitudes toward science. *Journal of College Science Teaching, 41*(4), 74-81.

*Wittwer, J., Nuckles, M., Landmann, N., & Renkl, A. (2010). Can tutors be supported in giving effective explanations? *Journal of Educational Psychology, 102*(1), 74-89. http://dx.doi.org/10.1037/a0016727

*Wozniak, B. M. (2012). *Effect of process-oriented guided-inquiry learning on non-majors biology students' understanding of biological classification* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1517251)

*Wright, J. M. (2008). The comparative effects of constructivist versus traditional teaching methods on the environmental literacy of postsecondary nonscience majors. *Bulletin of Science, Technology & Society, 28*(4), 324-337. http://dx.doi.org/10.1177/0270467608319638