Studies in Higher Education

ISSN: 0307-5079 (Print) 1470-174X (Online) Journal homepage: https://www.tandfonline.com/loi/cshe20

High-impact teaching practices in higher education: a best evidence review

Calvin D. Smith & Chi Baik

To cite this article: Calvin D. Smith & Chi Baik (2019): High-impact teaching practices in higher education: a best evidence review, Studies in Higher Education, DOI: 10.1080/03075079.2019.1698539

To link to this article: https://doi.org/10.1080/03075079.2019.1698539

STUDIES IN HIGHER EDUCATION
https://doi.org/10.1080/03075079.2019.1698539

Routledge
Taylor & Francis Group

High-impact teaching practices in higher education: a best evidence review

Calvin D. Smith [a] and Chi Baik [b]

[a]Office of Medical Education, Faculty of Medicine, University of Queensland, Brisbane, Australia; [b]Centre for the Study of Higher Education, Melbourne University, Melbourne, Australia

ABSTRACT

In this paper, we report on an attempt to systematically discover reliable and high-quality evidence on teaching practices that have an impact on student learning. We adopted an innovative approach that was modelled on some of the practices of a systematic review. Papers that met certain quality criteria (population generalisability, ecological validity, measurement validity, logical clarity and design justification) were examined and a weighted-average score for each paper was calculated. These were then ranked to derive lists of best-evidenced and most impactful teaching and curriculum strategies. The results reveal that this method provides a reasonable approach to distilling the available literature into a concise representation of the most effective practices that are backed by good research design attributes. The results will be useful for curriculum designers, university leaders and policy makers, and have already informed policy at one Australian university.

KEYWORDS

Teaching practices; evidence based teaching; higher education; teaching research and evaluation; student outcomes; teaching methods

Introduction

There is an abundance of literature on 'good teaching' in higher education, including a growing number of widely-used texts and resources offering research-informed advice for teaching in higher education (e.g. Biggs and Tang 2011; Laurillard 2001; Ramsden 2003).

However, many teachers base their teaching practices on tradition, the advice of experienced practitioners, personal experience gained through trial and error, 'ideology, faddism, marketing or politics' (Groccia and Buskist 2011, 5). Further, much of the research on teaching practices is conducted within particular disciplines on disciplinary cohorts of students; this is understandable because it is likely that much SOLT (scholarship of learning and teaching) activity and research on teaching practices is based in disciplines and so is conducted on samples of convenience within those disciplines – a threat to the external validity or generalisability of the findings of those studies (Leppink 2019, 8–10, 17). This is not to say that there are no studies across whole institutions or multiple disciplines, just that these are the minority (Cook and Hatala 2015).

A corollary of much SOLT research being based on discipline-based, convenience samples, is that many such studies have relatively small sample Ns – a is a threat to the reliability of estimates of effects (Leppink 2019, 55–56).

Another quality issue in research on teaching practices is the wide variety of both research questions and methods used (Bearman et al. 2012, 629; Ling and Ling 2020). There is an observable pattern that many studies are based on surveys, focus-groups and interviews with

CONTACT Calvin D. Smith calvin.smith@uq.edu.au Office of Medical Education, Faculty of Medicine, University of Queensland, Level 3, Mayne Medical Building, 288 Herston Rd., Herston, Brisbane, QLD 4006 Australia

students, which are often based on qualitative, interpretivist research designs; even those that are quantitative and objectivist in underlying philosophy may not use *objective* measures of outcomes but instead rely on study participants' self-reported outcomes; compounding this is that most studies deriving from the growing interest in SOLT are typically not *experimental* in design, and are most likely to be correlational rather than *comparative* (for a full review of this diversity see Ling and Ling 2020). Yet another threat to validity and reliability is that studies are often conducted as a once-off design, rather than being longitudinal in character (Leppink 2019, 54).

Again, this is not to say that there are no positivistic, quantitative, comparative, or longitudinal, studies of teaching practice effectiveness, that use objective outcomes measures, and an experimental design, only that these are in the minority.

Finally, higher education is a context in which it is less common, certainly than it is in the sciences, for communities of researchers to build on the work of others, or to follow a common line of inquiry into problems where the definition of concepts, and indeed of the problems themselves, is shared or agreed upon (Bearman et al. 2012). Though there are some concepts and constructs that are shared within the field of higher education studies, sustained, focused, and cumulative programmes of research are relatively rare. There are some exceptions – constructs that have had a sustained run – for instance, good examples of sustained inquiry and scholarship on a theme maybe the idea of students' approaches to learning (Biggs 1988; Entwistle, Hanley, and Hounsell 1979; Marton and Säljö 1976) or threshold concepts (Land, Meyer, and Flanagan 2016; Meyer and Land 2005).

Given the current and emerging challenges facing the higher education sector with the rapid growth in student participation and an increasingly diverse student cohort, there is a need, today more than ever before, for a focus on sound evidence-based teaching – that is, teaching practices that are most likely to lead to effective student learning outcomes (James et al. 2015). Evidence is not the only thing needed to secure quality; according to Hattie (Hattie 2015, 89)

> [f]aculty need to go beyond merely collecting data, creating reports, and asking students to fill in surveys, but to become excellent interpreters of evidence about their impact.

For instance, the measurement and reporting of effect sizes for studies of the impact of different teaching and learning strategies is relatively uncommon. Justifying a motivation to take effect sizes into account in appraising literature, Hattie observes that

> [o]ne of the surprising findings was that across the interventions that are commonly claimed to enhance student learning—nearly all of them have a positive impact on student learning. That is, almost everything works! But herein lies the greatest problem in education—every method seems to work relative to not implementing that method. … What we should be asking, instead, is the magnitude of the learning improvement—Some interventions have dramatically higher impacts than others. (Hattie 2015, 81)

This paper presents the findings of a systematic approach to the review of studies investigating the effects of different curriculum design and teaching practices on learning outcomes. The aim was to identify high-impact teaching practices – that is, practices that had the greatest effect on student learning – that were also generalisable, well-conceived, and tested in well-designed studies. The aim was to establish the warrant for policy recommendations for teaching practice and curriculum design.

The study

Method

Search and selection method

The search strategy used employed the *Web of Science* reference database and search engine; the syntax is listed in Table 1. (The full listing of journals searched is in Table 2.)

Table 1. Listing of syntaxes used for the searches in Web of Science database interface.

Search Number	Yield	Syntax[a]
1	8	SO=(… list of journal titles …) and TS=(higher education measure* teach* curricul* learning outcome*)
2	24	TS=(higher education measure* teach* achievement)
3	5	TS=(SEEQ)
4	5	TS=(SEEQ) including INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH and REVIEW OF EDUCATIONAL RESEARCH from this point on
5	349	TS=(effective teaching)
6	2	TS=(effective teaching attainment)
7	50	TS=(effective teaching outcome*)
		Other search strategies and strings
8	350 approx	searching in *Teaching Psychology* for the string 'effective teaching'
9	119	effective learning experiment university 'effect size' in *Learning And Instruction* the journal
10	43	effective learning experiment university 'effect size' in *International Journal of Science and Mathematics Education*
11	3	effective teaching learning experiment university 'effect size' in *Journal of Political Science Education*
12	0	effective teaching learning experiment university 'effect size' in *Journal of Social Science Education*
13	7	effective teaching in *Journal of Social Science Education*
14	407	'effective teaching learning experiment university effect size' in the *Int'l Journal Of Medical Education* (UK) – this search seemed to target beyond medical education
15	15	pubid(266697) AND (effective teaching) AND learning AND (effect size) AND university using ProQuest where pubid = *int'l journal medical education*
16	194	pubid(42593) AND (effective teaching) AND learning AND (effect size) AND university using ProQuest where pubid = *BMC Medical Education*

[a]Syntax for searches 1 through 7 is complete for search 1 in the Table; after that only the 'TS=' (which is the term search syntax) is included, as the journal list was the same for each search. Journals searched were the same except for the addition of two of the journals listed in Table 2 for searches 4–7.

An initial selection of papers was made by Author 1 (CS), based on *prima facie* evidence in the title and abstract that the paper would meet minimum criteria for inclusion. These minimum criteria included:

- that the paper was about a teaching and learning practice;
- that it reported on the appraisal/evaluation of that practice;
- that the appraisal of the practice was based on its impact on either student learning or student experience; and
- that the paper estimated an *effect size* either by comparison (that is to say it compared the effectiveness of a teaching practice on those who did not experience the practice with those who did), or by *estimating the impact* of the practice on learning or experiential outcomes, or by calculating the *correlation* of the practice with outcomes.

The papers that were selected by this method were then reviewed in more details (by Author 1). An initial review of abstracts generated a pilot list of about 50 papers, which the authors examined and the methodology slightly amended as a consequence. The amendments to the methodology related to the allocation of weighting and points to the inclusion criteria (see Table 3). Weightings were determined according to the purposes of the research and reflected these. Thus, ecological validity was important (weighting for this criterion was 1, and within it, the standards expressed a ranked preference for work done on/in naturalistic settings), but population generalisability and the validity of measurement were more important (weighting of 2, overall, with standards within each afforded refined but rank-ordered sub-scores). The criterion that we weighted (valued) most was clarity in the design and the logic and purposiveness of the study. Again, within that criterion, there were ranked standards. In this way, the weighting scheme embodied the *values* that underpinned the study – a search for high-quality, generalisable findings, derived from well-designed and coherently-reported studies (see Table 3).

Table 2. Listing of journals used in the searches.

CANADIAN JOURNAL OF HIGHER EDUCATION	INNOVATIONS IN HIGHER EDUCATION TEACHING AND LEARNING
CORE CONCEPTS IN HIGHER EDUCATION	HIGHER EDUCATION RESEARCH DEVELOPMENT
DEVELOPING CULTURAL CAPABILITY IN INTERNATIONAL HIGHER EDUCATION A NARRATIVE INQUIRY	HIGHER EDUCATION QUARTERLY
DIALOGUE BETWEEN HIGHER EDUCATION RESEARCH AND PRACTICE	KEY ISSUES IN HIGHER EDUCATION
EFFECTIVE LEARNING AND TEACHING IN HIGHER EDUCATION	NEW APPROACHES TO PROBLEM BASED LEARNING
ENHANCING QUALITY IN HIGHER EDUCATION INTERNATIONAL PERSPECTIVES	REVITALISING YOUR PRACTICE IN HIGHER EDUCATION
EVALUATION OF ONLINE HIGHER EDUCATION LEARNING INTERACTION AND TECHNOLOGY	PEER REVIEW OF LEARNING AND TEACHING IN HIGHER EDUCATION INTERNATIONAL PERSPECTIVES
EXCELLENCE IN HIGHER EDUCATION	QUALITY IN HIGHER EDUCATION
EXPLORING LEARNING TEACHING IN HIGHER EDUCATION	RESEARCH IN HIGHER EDUCATION
HIGHER EDUCATION	REVIEW OF EDUCATIONAL RESEARCH[a]
HIGHER EDUCATION AND CIVIC ENGAGEMENT COMPARATIVE PERSPECTIVES	REVIEW OF HIGHER EDUCATION
HIGHER EDUCATION AND COMMUNITY BASED RESEARCH CREATING A GLOBAL VISION	SERVICE LEARNING IN HIGHER EDUCATION CRITICAL ISSUES AND DIRECTIONS
HIGHER EDUCATION AND DEMOCRACY ESSAYS ON SERVICE LEARNING AND CIVIC ENGAGEMENT	STUDIES IN HIGHER EDUCATION
HIGHER EDUCATION FOR A KNOWLEDGE SOCIETY	TEACHING FOR LEARNING AND LEARNING FOR TEACHING PEER REVIEW OF TEACHING IN HIGHER EDUCATION
HIGHER EDUCATION HANDBOOK OF THEORY AND RESEARCH	TEACHING IN HIGHER EDUCATION
HIGHER EDUCATION HANDBOOK OF THEORY AND RESEARCH VOL 25	THEORY AND METHOD IN HIGHER EDUCATION RESEARCH
HIGHER EDUCATION HANDBOOK OF THEORY AND RESEARCH VOL 26	THEORY AND METHOD IN HIGHER EDUCATION RESEARCH I
HIGHER EDUCATION HANDBOOK OF THEORY AND RESEARCH VOL 29	THEORY AND METHOD IN HIGHER EDUCATION RESEARCH II
JOURNAL OF CONTINUING HIGHER EDUCATION	ACTIVE LEARNING IN HIGHER EDUCATION
JOURNAL OF HIGHER EDUCATION	ASSESSMENT EVALUATION IN HIGHER EDUCATION
ISSUES IN HIGHER EDUCATION	LEARNING AND INSTRUCTION
INTERNET AND HIGHER EDUCATION	APPLIED COGNITIVE PSYCHOLOGY
INTERNATIONAL JOURNAL OF EDUCATIONAL TECHNOLOGY IN HIGHER EDUCATION	JOURNAL OF EDUCATIONAL PSYCHOLOGY
INTERNATIONAL JOURNAL OF EDUCATIONAL RESEARCH[a]	BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY
INTERNATIONAL PERSPECTIVES ON HIGHER EDUCATION RESEARCH	COGNITIVE SCIENCE
INTERNATIONAL STUDIES IN HIGHER EDUCATION	MIND MODELLING A COGNITIVE SCIENCE APPROACH TO REASONING LEARNING AND DISCOVERY
INTEGRATING PRACTICE BASED EXPERIENCES INTO HIGHER EDUCATION	TEACHING OF PSYCHOLOGY

[a]Those entries marked with were added in searches 4 through 7.

The full set of potential inclusions was then reviewed (by Author 1) and a final set of (seventy-eight) papers was selected. The 78 papers provided information on 96 studies that met the inclusion criteria.

Evaluation method

Included papers were appraised and scored on four criteria. The criteria for the evaluation of the papers are listed in Table 3. The criteria were weighted, and along with the effect size (or equivalent) a formula was used to calculate a final score for each paper.

For each study, a weighted mean was calculated that combined Cohen's *d*, the score that the study achieved in each of the dimensions or criteria, and the weighting that was applied to each criterion. Where papers reported Cohen's *d* we incorporated this calculation directly into the weighted mean formula. Where findings were reported as Pearson's *r* or *r*-squared, or *F* statistics, a Cohen's *d* equivalent was calculated, and this calculated value was incorporated into the weighted mean calculation.

As a consequence, each study achieved a final weighted main score that took into account the paper's quality, in a scheme that weighted criteria deferentially according to decisions made by

Table 3. Criteria and weighting.

Criteria and standards for point allocation	Weighting
Population Generalizability Rating: 1 = one discipline/one university; 2 = one university & multiple disciplines/multiple universities & one discipline; 3 = multiple universities and multi-disciplines (one-off or multi times)	2
Measurement Validity rating: 1 = none/single items; 2 = Principal Components Analysis /Exploratory Factor Analysis (reporting factor scores or alphas only); 3 = Confirmatory Factor Analysis /Structural Equation Modelling	2
Ecological Validity rating: 1 = Lab experiment; 2 = Lab experiment with naturalistic material/simulated setting; 3 = naturalistic (naturally occurring) setting and materials	1
Logic of inquiry clarity: 1 = data dredging, unclear exposition of purpose, no link to theory; 2 = links to theory but exposition of purpose, design, measures, test is unclear; 3 = clear theoretical links, design, argument, variables and tests	3

the research team, and which suited the purposes of the study. For instance, studies that were conducted with cohorts of students from *multiple* disciplines scored more highly on *generalisability* than studies that we conducted the students from *single* disciplines; studies that were experimental and conducted *in laboratories* scored less highly on *ecological validity* than studies that had good designs but also good *ecological validity*.

Using this scheme, our ratings valued most highly *generalisable findings*, based on *well-measured* data/constructs, collected in studies conducted in *naturalistic settings*, with a very clear *purpose and logic* of inquiry. Further, the ratings scheme valued logic of inquiry and clarity of purpose, above measurement validity and generalizability, and all of these above ecological validity.

Procedure

Each paper's abstract was examined and the details compared with the criteria for inclusion. The criteria for inclusion were:

- the paper reported on a study that was experimental, quasi-experimental, or correlational (including ANOVA, regression, structural equation modelling, path-analysis);
- the paper reported finding/s that potentially could inform university teachers' and curriculum designers' practices;
- the paper reported a design that was comparative – that it is it explored the question of the impact on students' outcomes of the educational practice being studied, either compared with a null impact, or compared with an alternative or control group or practice;
- the paper reported a design the dependent variables of which included learning outcomes, not just satisfaction measures.

If the abstract immediately justified inclusion *prima facie*, the paper was retained for further consideration and possible summarisation. If the abstract indicated that the paper did not meet the criteria for inclusion it was excluded. If the abstract did not make the decision to exclude unequivocally possible, then the paper was included for further analysis. Included papers were then examined in more detail. Those that did not meet criteria for inclusion were excluded. Those that met the criteria for inclusion were retained and explored in more detail in a further step. In that final step, some papers were discarded because they did not measure up to the promise of their abstracts, or, for some other reason that became apparent upon closer inspection, they did not meet the criteria for inclusion.

Those that were retained were summarised in a spread-sheet. There were seventy-eight (78) such papers retained and summarised.

Key summary information

Along with the full citation details, the following data were extracted for each paper:

- analysis sample size
- sample characteristics and source/s
- discipline group [1]
  - 0 = not reported;
  - 1 = STEM;
  - 2 = HASS;
  - 3 = Mixed
- theoretical focus
- inputs studied
- mediators studied
- controlling for
- outcomes – higher-order grouping (a broad label for the outcome clusters)
- outcomes studied (more detailed description of outcomes studied)
- methods and statistics used.

Effect sizes

Where the paper reported effect size/s as Cohen's *d*, we used that in the calculation of the weighted mean (see below). Where the paper reported statistics that could be interpreted as or converted to effect size equivalents (*eta*-squared, *R*-squared, Hedges *g*, omega-squared, *z*-scores, *chi*-squared, Pearson's *r*, or *F* for an ANOVA) Cohen's *d* values were calculated using formulae summarised in Beasley (2016) and Field (2009).

Weighted average of quality and effect size used to rank papers

A weighted average for each paper was calculated that took into account the quality indicator scores and the size of the effect reported.

The weighted average included the effect size *d* x 100, and was derived by the following formula:

$$\text{Avg}^{w} = (w_g.g + w_{mv}.mv + w_{ev}.ev + w_{lc}.lc + 100.d)/5 \quad (1)$$

where w_g is the weight given to generalisability; g is the generalisability score; w_{mv} is the weight applied to measurement validity; mv is the measurement validity score; w_{ev} is the weight applied to ecological validity; ev is the ecological validity score; w_{lc} is the weight applied to logical clarity; lc is the logical clarity score and d is the Cohen's *d*.

The papers were then sorted by their final weighted average and the results reported in three broad bands (tertiles):

- top 33% (Avg^{w} = 49.05–20.42; d = 2.3–0.795);
- middle 33% (Avg^{w} = 20.23–13.11; d = 0.81–0.46); and
- bottom 33% (Avg^{w} = 12.80–1.26; d = 0.46–0.22).

For each paper, a summary of the implications for teaching practice or curriculum design was made. These summaries were then grouped or categorised. Categories were created as short 1–3 word group labels that pointed to the focus of the paper, in terms of educational theory, teaching practice or curriculum design (see Table 4).

Table 4. List of short category labels and *N* of studies included for each.

Short label	*N*	Short label	*N*	Short label	*N*	Short label	*N*
Structure	11	Relationship (student-teacher)	2	Blended learning – high independence	1	Intensives	1
Alignment	5	Spaced revision	2	Cognitive and meta-cog. prompting; Schema	1	Learning community	1
Active learning	3	Teacher skills	2	Cognitive and meta-cognitive prompting; Writing	1	Narration + animation in Multi-media	1
Collaborative learning	3	Teacher skills – Elocution	2	Cognitive challenge	1	PBL (problem-based learning) – for knowledge *versus* skill development	1
Inquiry-based learning	3	Value	2	Collaborative assessment	1	Peer assessment	1
Interaction/dialogue	3	Variation theory	2	Collaborative learning + Expertise	1	Peer-tutoring will develop you meta-cognition	1
PBL (problem-based learning)	3	Writing	2	Dialogue (Blended L)	1	Retrieval practice	1
Relationship and structure	3	Alignment and application of knowledge	1	Dialogue + Application	1	Student approaches to learning (SAL)	1
Cognitive and meta-cog. prompting	2	Alignment and Professional dev't of teachers	1	Gaming (Effect is negative)	1	Small group teaching – optimum sizes	1
Episodic richness	2	Application of knowledge	1	Exam Practice	1	Structure and cognitive load	1
Expertise	2	Application of knowledge and Simulation	1	Facilitation of learning	1	Stimulate interest	1
Inter-teaching/ flipping	2	Assessment literacy – meta-cognition	1	'Teaching Quality' (CEQ measure) is not useful	1	Teaching peers	1
Meta-cognitive training	2	Awareness of learning/ progress	1	Heuristic examples/ modelling	1	Workload	1
Randomised practice	2	Blended learning	1	Independent learning	1	Total	94

Results

The main findings can be described and documented in a variety of ways. The first, most obvious approach, is a simple sort on the size of the final weighted average.

These results can be further summarised by counting the number of times, and in which tertile, each encoding occurs. In Table 5, the data are reduced further into counts of category encodings, across all papers, *within* the tertiles bands in which the papers sit. The data are then sorted hierarchically (first by 3rd tertile, then 2nd, then 1st); on this basis, a more succinct picture emerges. This picture tells the story of the more frequently-studied constructs/practices and the spread of papers reporting them across the quality ratings. The definitions of each of these codes are presented in rank order in Table 6.

Translating the findings to inform practice

Table 7 provides an example translation of these findings into a higher level of abstraction, useful for institutional teaching and learning policy. In it, the categories/codes derived in the review (centre column) are further grouped into cognate collections (left-hand column), and an explication of their defining characteristics is provided (right-hand column).

It is important to note that the cognate groupings are not discrete but often overlap. For example, 'experience' often involves 'application' but of a particular practical or 'real-world' kind. Similarly, there is an overlap between 'inquiry' and 'challenge'.

Table 5. Final sort – by summary of encodings (some findings multiple-coded) of 1st, 2nd and 3rd tertile representing the number of times the 'method' occurred in each tertile.

Code	Top 33%	Middle 33%	Low 33%	Unclass-ified	Total encodings
Structure	8	6	1		15
Alignment	2	3	2		7
Collaborative learning	2	1	1		4
PBL	2		2		4
Inter-teaching/flipping	2				2
Relationship	1	4			5
Active learning	1	2			3
Application of knowledge	1	1	2	1	5
Expertise	1	1	1	1	4
Episodic richness	1	1			2
Meta-cognitive training	1	1			2
Randomised practice	1	1			2
Inquiry-based learning	1		2		3
Awareness of learning/progress	1				1
Gaming (N.B. the effect is negative)	1				1
Heuristic examples/modelling	1				1
Independent learning	1				1
Intensives	1				1
Narration + animation in Multi-media	1				1
Peer-tutoring will develop you meta-cognition	1				1
Retrieval practice	1				1
Stimulate interest	1				1

Discussion

The purpose and intent of this study were to provide any higher education institution an evidence-based approach to justify, recommend and formulate policy on teaching practice across the institution. This review was not a *systematic review* in the sense of Cochran or Campbell approaches, however it was conducted systematically and was pragmatically designed with the overall purposes of the study in mind. Cochran-like systematic reviews are not viable in higher education because of variability in the meaning of terms, variety in research methods, and dearth of experimental designs; a consequence is that the searches that strictly adopt search-term rules, can miss important papers that should be included in a review. This was argued in the paper by Bearman et al. (2012). Even when a pragmatic approach is adopted, being systematic is not straightforward. Nonetheless, it is possible and the results can reward the effort.

The literature contains some clearly-written reports of well-designed studies that give some indication of robust and powerful teaching and curriculum design strategies that have been tested and shown to be effective. Whilst there are many individual studies that identify specific high-impact practices, the quality across this collection is variable. The significance of this study is that it identifies a sub-set of high-quality papers using *a priori* criteria and standards, and pulls together the findings across that sub-set of papers. In this way, it is the *quality of the research* selected, rather than the *narrative focus of the review* (e.g. a search for evidence on a particular practice) that determines the outcome of the study. Thus, although the quality criteria and standards were set by the authors, rather than standing independently of the study, it was the quality criteria themselves that generated the ranking of the final selected papers and their respective findings.

This study adds to our knowledge of effective teaching practices in two ways: first it selects studies based on quality and generalisability criteria and second, it ranks findings on the basis of quality criteria combined with effect size. This gives us a different window onto, or view of, the literature on effective practices for facilitating learning than is usually provided through either systematic review or narrative review. Since the criteria themselves were generated to serve the purpose of the study, which was to provide a defensible, best-evidence-based, summary of generally applicable, highly effective practices thus the results indicate the kinds of practices that can be relied upon across a range of contexts and applications in higher education.

Table 6. Final top tertile collection of constructs and definitions.

Construct name/label	Definition or explanation
Structure	The provision to students of clarity in both the representation of conceptual content, relationships between ideas and so forth, and in the step-wise development of learning towards a well-defined end and purpose. In SEEQ there are two related dimensions: Teacher's Preparation; Organisation of the Course and Clarity and Understandableness; e.g. 'The instructor gave clear explanations'; 'the instructor interprets abstract ideas and theories clearly'. (Marsh 1982, 80)
Alignment	Biggs' Constructive alignment – designing into the curriculum activities (in and out-of-class) that support the achievement of the goals and that mirror the assessments. (Biggs 1996)
Collaborative learning	Students learning in teams (also: team-based learning).
PBL	Problem-based learning
Inter-teaching/flipping	The practice of giving trigger materials/content to students in advance of meeting with them in the lecture time and using the lecture time to discuss, explore, interrogate or extend the ideas in the trigger materials.
Relationship	A SEEQ dimension: Teacher's Availability and Helpfulness; also operationalisation of 'rapport' through such attributes as: eager to help students; role model; cares about students; encourages me to succeed; enthusiastic; compassionate; reliable; makes class enjoyable; receptive; understanding; thoughtful; communicates well; spends extra time going over a concept if students need it; friendly; considerate; approachable. (Marsh 1982, 80)
Active learning	Students actively engaged in their learning, whether stimulated to do so by the teacher or the curriculum, or by dint of their personal approach to their studies.
Application of Knowledge	Having students apply what they have learned (whether conceptual or psycho-motor) in increasingly authentic real-world-like settings and/or to authentic real-world-like problems.
Expertise	The use of expert teachers (lecturers or practitioners) rather than peer teachers or other non-experts.
Episodic richness	Giving students examples of the application or instantiation of concepts via real-life, detailed, specific and anecdotal accounts.
Meta-cognitive training	Giving students training in meta-cognition.
Randomised practice	Randomising the occurrence of revision/practise of a concept/skill.
Inquiry-based learning	Involving students in the inductive/research/inquiry process, investigating and solving disciplinary problems.
Awareness of learning/progress	Giving students the capacity to recognise how they have developed/learned, that they have developed/learned.
Gaming (N.B. the effect is negative)	The use of virtual reality environments often featuring pseudo-identities (avatars) and sometimes reward and challenge outcomes as ways to engage students in the problems/content to be learned.
Heuristic examples/modelling	Allowing students to witness the working-through of a problem or reasoning so the thinking, planning or approach are made explicit (see also meta-cognition).
Independent learning	Students self-instructing (as compared with being instructed by an expert).
Intensives	Curriculum design in which instead of semester-long engagement with a few hours per-week, there is 1–3 weeks of full-time engagement with students.
Narration + animation in Multi-media	As opposed to text only, the use of spoken word (narration) along with animation to describe and illustrate concepts.
Peer-tutoring will develop you meta-cognition	Having student teach concepts or techniques to other students.
Retrieval practice	Opportunities to retrieve from memory ideas already learned; the active, cue-driven process of reconstructing knowledge
Stimulate interest	A dimension in SEEQ: Teacher's Stimulation of Interest in the Course and Its Subject Matter 'The instructor puts material across in an interesting way'; 'the instructor gets students interested in the subject'; 'it was easy to remain attentive'; 'the teacher stimulated intellectual curiosity' (Marsh 1982, 80).

In summary these results they tell us that quality learning environments are those that:

- provision well-structured representations of disciplinary knowledge/concepts, in well-structured and clearly-planned subject/programme contexts;
- are intellectually challenging;
- take into account students' goals, and make clear the relevance of what they are learning;
- deploy expert teachers who build rapport with students;
- facilitate application/practice opportunities in authentic or simulated practice situations;

Table 7. Example translation to policy-level statement based on findings on High impact curriculum design and teaching practices

'High impact' practices focus on:	Categories/codes derived from the review	Defining characteristics
1. **Clarity**	Structure of content representations Alignment Expertise Relationship	Making the structure of knowledge and the progression of learning clear for students. There are three levels here: (a) Curriculum design – clear objectives and alignment between objectives, activities and assessments; clear organisation of topics within a subject ('structure of disciplinary content') (b) Design of lectures/classes – clear planning & organisation of content & activities (c) Delivery – clear explanations and content structuring by experts
2. Inquiry	PBL (problem-based learning) CBL (case-based learning) Active learning Inquiry-based learning	Using approaches/methods that have the goal of fostering questioning, problem solving, investigating and testing. Sometimes referred to as 'inquiry-based learning' or 'active learning'. Examples of common pedagogical approaches include PBL, CBL.
3. Application	Active learning Application of Knowledge Inter-teaching ('flipping' classes) PBL (problem-based learning) CBL (case-based learning)	Involving students in exercises/activities to apply knowledge and understanding.
4. Experience	Active learning PBL (problem-based learning) Episodic richness	A particular kind of application that is about practical, experiential learning, sometimes referred to as 'authentic learning' or 'real world' practice (e.g. WIL)
5. Challenge	Stimulate interest Inquiry-based learning	Stimulating interest and fostering deep cognitive engagement. Sometimes mentioned in relation to 'inquiry-based learning' and 'PBL',
6. Relevance	Value (to students) Episodic richness Stimulate interest	Helping students to see the value/purpose in what they are learning. There are two levels to this: (1) Pedagogical approaches – experiential learning, CBL, PBL (2) Teacher's delivery – e.g. using authentic examples of disciplinary ideas/constructs to students
7. Interaction and relationships	Collaborative learning Interaction/dialogue Student – teacher relationship Collaborative assessment Peer-tutoring	Enabling and facilitating peer interaction and learning in a social context; Fostering positive interaction between students and teachers
8. Consolidation	Randomised practice Exam practice Retrieval practice Structure of content representations; relationships between ideas	Providing appropriate kinds of retrieval practice and revision, in which material to be learned is 'retrieved' during study sessions subsequent to the first session in which material is learned. Consolidating understanding and correcting misconceptions
9. **Self-regulation**	Metacognitive training Modelling Awareness of learning/progress Independent learning	Facilitating students' self-assessment, managing their own learning (e.g. planning, organisation, monitoring, remedial action, reviewing), learning how to learn and thinking about how they come to learn.

- give students opportunities to engage in inductive/exploratory/dialogic learning;
- give students opportunities to interact and work with peers; and
- take into account students' own role and agency in their learning, encouraging them to engage the meta-cognitive processes that ensure long-term encoding and retrieval, and which develop meta-cognition skills that underpin self-assessment, self-monitoring and management of their own learning.

There are, as is often the case, some caveats: First, the review strategy selected papers that had a comparative design for which an effect size was calculated. That effect size is typically, though not always, based on a comparison of means of an experimental and control group. One problem with this is that in some cases, the control group was a 'do nothing' condition (e.g. participants in the control group did anagrams at the same time that the student in the experimental group engaged in the experimental condition). In cases such as these, the effect size can be quite large, because the comparison group has had no intervention of an *educationally relevant* kind. The problem we are often trying to address is more practical and of the form: 'Is this educational practice better than some other educational practice?'(e.g. is small group work better than lecturing (on this note see also Hattie 2015, 81)).

Second, different *weightings* for the quality variables would have created different final results (though these differences might have been only at the margins of the lists not in the overall pattern of results). Thus, the results are influenced heavily by the way the weightings privilege some study attributes over others. In the context, this is not a weakness because the authors wanted to find papers that met these criteria, but it is important to bear these weightings in mind when interpreting the results.

As always, the overarching concern in selecting or encouraging the selection of a teaching strategy or curriculum design feature should be the 'fit' of the method with the intended outcomes. To be able to provide advice to individual teachers from a study such as this review, it should be conducted with the specific intended outcomes (and perhaps several other factors, such as student presage factors) in mind; this review was conducted as a search for studies that show impact in *any class* of learning outcome. This is extremely important; any work that focuses on effect sizes but does not take into account the context, or recognise different conceptualisations of independent and dependent variables, risks arriving at invalid conclusions. Hattie's recent work comparing effect sizes is a case in point; although the independents appear to be all of a kind, there is no clear effort to match dependents, the context, or the purposes of the studies compared (Hattie 2008, 2015; Hattie and Yates 2014). Therefore, ranking a multitude of apparently alike studies by their effect sizes does not give the information practitioners need to decide if the top or best-ranked methods are right for them, *in their context*, *with their purposes*.

Educational research and studies on learning outcomes cannot be interpreted or compared in a straightforward way. Much depends on the particular context in which the studies took place, and particularly in terms of whether a curriculum design feature or intervention will work or not, a lot depends on the individual knowledge and skill of the teachers designing the tasks and facilitating learning.

What is 'effective' teaching?

The literature on effective teaching, as already stated, contains many 'advisory' texts, each of which derives the justifications for their various advice points from a variety of sources. The question arises: are these points of advice equally well-supported in the research literature, when the quality of the research upon which they are based is taken into account? Using the approach adopted in this study, this question can be answered with some nuance. Putting the quality of research first (albeit, restricted to our way of defining *quality*), we are now in a position of saying which pieces of advice from various sources have strong support from the body of research in which they feature.

While there is no universally accepted definitions, broadly speaking, it is generally understood that effective teaching is that which leads effectively to students' learning. There is a large body of research over decades identifying common factors that are likely to contribute to successful outcomes for students (Biggs and Tang 2011; Chickering and Gamson 1987; Koljatic and Kuh 2001; Pascarella and Terenzini 1991). As Ramsden (2003) notes:

> A great deal is known about the characteristics of effective university teaching. It is undoubtedly a complicated matter; there is no indication of one 'best way,' but our understanding of its essential nature is both broad and deep. (88–89)

This review, however, sheds some light on what practices have been studied in a way that gives confidence in the findings, and have shown promising results in those studies. Adopting this approach, the lists below show summaries of principles expounded in a variety of texts, alongside an indicator for each as to whether it was supported through this review's findings.

Several prominent scholars have synthesised the body of research and have depicted effective teaching as being characterised by several core principles. For example, Chickering and Gamson's (1987), widely-cited 7 principles for good teaching based on 50 years of research into college teaching and learning, aims to provide board guidance for educators in undergraduate education:

1. Encourage contact between students and faculty	✓
2. Develop reciprocity and cooperation among students	✓
3. Encourage active learning	✓
4. Give prompt feedback	N.D.
5. Emphasise time on task	N.D.
6. Communicate high expectations	N.D.
7. Respect diverse talents and ways of learning	N.D.

Paul Ramsden's (2003) highly-cited six key principles of effective teaching emphasises good teaching in relation to the student experience:

1. Interest and explanation	✓
2. Concern and respect for student and student learning	✓
3. Appropriate assessment and feedback	N.D.
4. Clear goals and intellectual challenge	✓
5. Independence, control and engagement	✓
6. Learning from students	N.D.

Some researchers have focused on the perspectives and practices of 'exemplary' teachers themselves to characterise the core components of effective teaching, For example, Kember and McNaught's ten principles of effective teaching are based on interviews with 62 'award-winning' teachers (Kember and McNaught 2007), as is Duarte's Conceptions of Good Teaching (Duarte 2013) whilst Bain interviewed and studied the practices of over 40 exemplary teachers to understand what 'the best college teachers do' (Bain 2004).

More recently, there has been increased attention on the science of learning or 'how learning works' and the implications this has for teaching practice (Ambrose et al. 2010). Drawing on research in psychology, education and cognitive science, Ambrose et al. (2010) proposed seven principles of learning for 'smart teaching':

1. Student' prior knowledge can help or hinder learning	N.D.
2. How students organise knowledge influences how they learning and apply what they know	✓
3. Students' motivation determines, directs and sustains what they do to learn	N.D.
4. To develop mastery, students must acquire component skills. practice integrating them, and know when to apply what they have learned	✓
5. Goal-directed practice coupled with targeted feedback enhances the quality of students' learning	✓
6. Students' current level of development interacts with the social, emotional and intellectual climate of the course to impact learning	N.D.
7. To become self-directed learners, students must learn to monitor and adjust their approaches to learning.	✓

In addition, as mentioned earlier, there are thousands of small-scale studies investigating the connection between particular teaching methods or practices and student achievement. Hattie's (2008) synthesis of 800+ meta-analytic studies identified three key strategies that effective teachers employ for enhancing student achievement:

1. Communicate Clear learning intentions and criteria for success	✓
2. Use multiple teaching strategies that emphasise student perspectives in learning	N.D.
3. Seek feedback regarding the effectiveness of their teaching and provide feedback to students regarding the effectiveness of their learning	N.D.

The lists above indicate that there is fairly good-quality evidence for many of the practices recommended in a variety of core texts in higher education. The fact that some of the recommended practices were not supported in the corpus of materials we reviewed does not mean that those practices are not supported in the research literature – the evidence for them may have been missed in our search strategy, or the papers that report that evidence may not have met our criteria for inclusion. But what this study shows is that when quality criteria are used in the selection process, and when those criteria are used to score candidate papers in a nuanced (graded) way, and weighted, *along with effect sizes*, the papers that are identified give an evidence base that has certain characteristics that may further strengthen the justification for recommending the practices evaluated in those papers.

Note

1. We do not report results partitioned into discipline groupings in this paper.

Acknowledgements

We gratefully acknowledge the early advice of Professor Richard James of The University of Melbourne.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This review was funded by the Centre for the Study of Higher Education, The University of Melbourne.

Notes on contributors

Calvin D. Smith is Associate Professor, and Academic Lead for Professional Development, in the Office of Medical Education, in the Faculty of Medicine, at the University of Queensland, in Brisbane, Australia. His research focuses primarily on theorising the impact of work-integrated learning curriculum on integrative learning, lifelong learning and professional identity formation. He conducts research and writes on a variety of issues in higher education including curriculum, teaching practices, evaluation of teaching, scholarship of learning and teaching, and faculty development.

Chi Baik is an Associate Professor in higher education in the Centre for the Study of Higher Education at Melbourne University, in Australia. Her research examines factors affecting the experiences and educational outcomes of diverse students in contemporary higher education. She has led major studies that have contributed to informing institutional policies and practices including national projects on the first-year experience, student mental wellbeing, internationalisation of the curriculum, and the academic workforce and has conducted numerous studies on aspects of university teaching and learning including assessment of student learning, peer and collaborative learning, and interdisciplinary curriculum design.

ORCID

Calvin D. Smith http://orcid.org/0000-0002-3544-8337
Chi Baik http://orcid.org/0000-0002-0480-5522

References

Ambrose, S., M. W. Bridges, M. DiPietro, M. C. Lovett, M. K. Norman, and R. E. Mayer. 2010. *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco, CA: John Wiley & Sons.
Bain, K. 2004. *What the Best College Teachers Do*. Cambridge, MA: Harvard University Press.

Bearman, M., C. D. Smith, A. Carbone, S. Slade, C. Baik, M. Hughes-Warrington, and D. L. Neumann. 2012. "Systematic Review Methodology in Higher Education." *Higher Education Research and Development* 31 (5): 625–40. doi:10.1080/07294360.2012.702735.

Beasley, M. 2016. Effect Size Conversion.

Biggs, J. 1988. "Assessing Student Approaches to Learning." *Australian Psychologist* 23 (2): 197–206.

Biggs, J. 1996. "Enhancing Teaching Through Constructive Alignment." *Higher Education* 32 (3): 347–64.

Biggs, J., and C. Tang. 2011. *Teaching for Quality Learning at University*. 4th ed. Maidenhead: McGraw-Hill Education.

Chickering, A., and Z. Gamson. 1987. Seven Principles for Good Practice in Undergraduate Education. *AAHE BULLETIN*, March, 2–6. doi:10.1200/jco.2014.32.15_suppl.8018.

Cook, D. A., and R. Hatala. 2015. "Got Power? A Systematic Review of Sample Size Adequacy in Health Professions Education Research." *Advances in Health Sciences Education* 20 (1): 73–83. doi:10.1007/s10459-014-9509-5.

Duarte, F. P. 2013. "Conceptions of Good Teaching by Good Teachers: Case Studies from an Australian University." *Journal of University Teaching & Learning Practice* 10 (1): 1–15.

Entwistle, N., M. Hanley, and D. Hounsell. 1979. "Identifying Distinctive Approaches to Studying." *Higher Education* 8(4): 365–80.

Field, A. 2009. *Discovering Statistics Using SPSS*. 3rd ed. London: Sage Publications.

Groccia, J., and W. Buskist. 2011. Need for Evidence-Based Teaching. In *New Directions for Teaching and Learning* (pp. 5–11). doi:10.1002/tl.

Hattie, J. 2008. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Hattie, J. 2015. "The Applicability of Visible Learning to Higher Education." *Scholarship of Teaching and Learning in Psychology* 1 (1): 79–91. doi:10.1037/stl0000021

Hattie, J., and G. Yates. 2014. *Visible Learning and the Science of How We Learn*. London: Routledge.

James, R., C. Baik, V. Millar, R. Naylor, E. Bexley, G. Kennedy, and S. Booth. 2015. *Advancing the Quality and Status of Teaching in Australian Higher Education: Ideas for Enhanced Professional Recognition for Teaching and Teachers*. Sydney: Department of Education and Training, Australian Government.

Kember, D., and C. McNaught. 2007. *Enhancing University Teaching: Lessons from Research into Award Winning Teachers*. London: Routledge.

Koljatic, M., and G. D. Kuh. 2001. "A Longitudinal Assessment of College Student Engagement in Good Practices in Undergraduate Education." *Higher Education* 42 (3): 351–71. doi:10.1023/A:1017993113390

Land, R., J. H. F. Meyer, and M. T. Flanagan, eds. 2016. *Threshold Concepts in Practice*. Rotterdam: Sense.

Laurillard, D. 2001. *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*. 2nd ed. London: Routledge.

Leppink, J. 2019. *Statistical Methods for Experimental Research in Education and Psychology*. doi:10.1007/978-3-030-21241-4_2.

Ling, L., and P. Ling. 2020. *Emerging Methods and Paradigms in Scholarship and Education Research*. doi:10.4018/978-1-7998-1001-8.

Marsh, H. W. 1982. "SEEQ: A Reliable, Valid, and Useful Instrument for Collecting Students' Evaluations of University Teaching." *British Journal of Educational Psychology* 52 (1): 77–95. doi:10.1111/j.2044-8279.1982.tb02505.x

Marton, F., and R. Säljö. 1976. "On Qualitative Differences in Learning: I. Outcome and Process." *British Journal of Educational Psychology* 46(1): 4–11.

Meyer, J. H. F., and R. Land. 2005. "Threshold Concepts and Troublesome Knowledge (2): Epistemological Considerations and a Conceptual Framework for Teaching and Learning." *Higher Education* 49 (3): 373–88. doi:10.1007/s10734-004-6779-5

Pascarella, P., and E. Terenzini. 1991. *How College Affects Students*. San Francisco, CA: Jossey-Bass.

Ramsden, P. 2003. *Learning to Teach in Higher Education*. London: Routledge/Falmer.

Appendix. The final selection of 78 papers that supplied information on 96 studies

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2014). Strategies for Teaching Students to Think Critically: A Meta-Analysis. Review of Educational Research, (June 2015), 1–40. doi:10.3102/0034654314551063

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis. Review of Educational Research, 78(4), 1102–1134. doi:10.3102/0034654308326084

Baeten, M., Dochy, F., & Struyven, K. (2013). The effects of different learning environments on students' motivation for learning and their achievement. British Journal of Educational Psychology, 83(3), 484–501. doi:10.1111/j.2044-8279.2012.02076.x

Benton, S. L., & Cashin, W. E. (2010). Student Ratings of Teaching: A Summary of Research and Literature. Idea Paper # 50, 1–22. Retrieved from www.theideacenter.org

Berg, I. van den, Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. Studies in Higher Education, 31(3), 341–356. doi:10.1080/03075070600680836

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A Meta-Analysis of Three Types of Interaction Treatments in Distance Education. Review of Educational Research, 79(3), 1243–1289. doi:10.3102/0034654309333844

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. Learning and Instruction, 17(5), 564–577. doi:10.1016/j.learninstruc.2007.09.007

Catano, V. M., & Harvey, S. (2011). Student perception of teaching effectiveness: development and validation of the Evaluation of Teaching Competencies Scale (ETCS). Assessment & Evaluation in Higher Education, 36(6), 701–717. doi:10.1080/02602938.2010.484879

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. Learning and Instruction, 29(2014), 153–170. doi:10.1016/j.learninstruc.2012.05.003

Dagyar, M., & Demirel, M. (2015). Effects of problem-based learning on academic achievement: A meta-analysis study. Egitim ve Bilim, 40(181), 139–174. doi:10.15390/EB.2015.4429

De Backer, L., Van Keer, H., & Valcke, M. (2014). Promoting university students' metacognitive regulation through peer learning: the potential of reciprocal peer tutoring. Higher Education, (2015), 469–486. doi:10.1007/s10734-014-9849-3

de Bruijn-Smolders, M., Timmers, C. F., Gawke, J. C. L., Schoonman, W., & Born, M. P. (2014). Effective self-regulatory processes in higher education: research findings and future directions. A systematic review. Studies in Higher Education, 5079(February 2015), 1–20. doi:10.1080/03075079.2014.915302

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. Learning and Instruction, 13(5), 533–568. doi:10.1016/S0959-4752(02)00025-7

Drabick, D. A. G., Weisberg, R., Paul, L., & Bubier, J. L. (2007). Keeping it short and sweet: Brief, ungraded writing assignments facilitate learning. Teaching of Psychology, 34(3), 172–176. doi:10.1080/00986280701498558

Eaton, S. B., & Bean, J. P. (1995). An Approach/Avoidance Behavioral Model of College Student Attrition. Research in Higher Education, 36(6), 617–645.

Elicker, J. D., & McConnell, N. L. (2011). Interactive Learning in the Classroom: Is Student Response Method Related to Performance? Teaching of Psychology, 38(3), 147–150. doi:10.1177/0098628311411789

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. Research in Higher Education, 30(6), 583–645. doi:10.1007/BF00992392

Feldman, K. a. (1996). Identifying exemplary teaching: Using data from course and teacher evaluations. New Directions for Teaching and Learning, 1996(65), 41–50. doi:10.1002/tl.37219966509

Fenollar, P., Román, S., & Cuestas, P. J. (2007). University students' academic performance: An integrative conceptual framework and empirical analysis. British Journal of Educational Psychology, 77(4), 873–891. doi:10.1348/000709907X189118

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of Problem-Based Leaming : A Meta-Analysis From the Angle of Assessment. Review of Educational Research, 75(1), 27–61. doi:10.2307/3516079

Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: evidence of learning in a controlled inclass study of undergraduate students. Journal of College Science Teaching, 43(3), 83–89. doi:10.1017/CBO9781107415324.004

Grimstad, K., & Grabe, M. (2004). Are Online Study Questions Beneficial? Teaching of Psychology, 31(2), 143–146. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ683175&site=ehost-live

Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. Learning and Instruction, 16 (2 SPEC. ISS.), 122–138. doi:10.1016/j.learninstruc.2006.02.001

Haak, D. C., HilleRisLambers, J., & Freeman, E. P. S. (2011). Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. Science, 332(3 June 2011), 1213–1216. doi:10.1017/CBO9781107415324.004

Hecimovich, M., & Volet, S. (2014). Simulated learning in musculoskeletal assessment and rehabilitation education: comparing the effect of a simulation-based learning activity with a peer-based learning activity. BMC Medical Education, 14, 253. doi:http://dx.doi.org/10.1186/s12909-014-0253-6

Heijltjes, A., van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. Learning and Instruction, 29(2014), 31–42. doi:10.1016/j.learninstruc.2013.07.003

Helsdingen, A. S., van Gog, T., & van Merrienboer, J. J. G. (2011). The effects of practice schedule on learning a complex judgment task. Learning and Instruction, 21(1), 126–136. doi:10.1016/j.learninstruc.2009.12.001

Herbert, D. M. B., & Burt, J. S. (2004). What do students remember? Episodic memory and the development of schematization. Applied Cognitive Psychology, 18(1), 77–88. doi:10.1002/acp.947

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. Learning and Instruction, 18(1), 54–65. doi:10.1016/j.learninstruc.2006.10.008

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. Learning and Instruction, 36(2015), 38–45. doi:10.1016/j.learninstruc.2014.11.001

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. Science, 331(6018), 772–775. doi:10.1017/CBO9781107415324.004

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: Does comparison help? Learning and Instruction, 23(1), 90–97. doi:10.1016/j.learninstruc.2012.07.004

Kopp, K. J., Britt, M. A., Millis, K., & Graesser, A. C. (2012). Improving the efficiency of dialogue in tutoring. Learning and Instruction, 22(5), 320–330. doi:10.1016/j.learninstruc.2011.12.002

Kühl, M., Wagner, R., Bauder, M., Fenik, Y., Riessen, R., Lammerding-Köppel, M., … Celebi, N. (2012). Student tutors for hands-on training in focused emergency echocardiography - a randomized controlled trial. BMC Medical Education, 12(1), 101–106. doi:10.1186/1472-6920-12-101

Larkin, H., & Richardson, B. (2013). Creating high challenge/high support academic environments through constructive alignment: student outcomes. Teaching in Higher Education, 18(2), 192–204. doi:10.1080/13562517.2012.696541

Lewis, J. S., & Harrison, M. A. (2011). Online delivery as a course adjunct promotes active learning and student success. Teaching of Psychology, 39(1), 72–76. doi:10.1177/0098628311430641

Lizzio, A., Wilson, K., & Simons, R. (2002). University Students ' Perceptions of the Learning Environment and Academic Outcomes : implications for theory and practice. Studies in Higher Education, 27(1), 37–41. doi:10.1080/0307507012009935

Lloyd, S. A., & Robertson, C. L. (2012). Screencast Tutorials Enhance Student Learning of Statistics. Teaching of Psychology, 39(1), 67–71. doi:10.1177/0098628311430640

Loes, C. N., Salisbury, M. H., & Pascarella, E. T. (2015). Student perceptions of effective instruction and the development of critical thinking: a replication and extension. Higher Education, 69(5), 823–838. doi:10.1007/s10734-014-9807-0

Mahling, M., Münch, A., Schenk, S., Volkert, S., Rein, A., Teichner, U., … Celebi, N. (2014). Basic life support is effectively taught in groups of three, five and eight medical students: a prospective, randomized study. BMC Medical Education, 14, 185. doi:10.1186/1472-6920-14-185

Marsh, H. W., & Ware Jr., J. E. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. Journal of Educational Psychology, 74(1), 126–134. doi:10.1037/0022-0663.74.1.126

Marsh, H. W., Hau, K.-T., Chung, C.-M., Siu, T. L. P., & Teresa, L. P. (1997). Students' evaluations of university teaching: Chinese version of the Students' Evaluations of Educational Quality Instrument. Journal of Educational Psychology, 89(3), 568–572. doi:10.1037/0022-0663.89.3.568

Masui, C., & Corte, E. (2005). Learning to reflect and to attribute constructively as basic components of self-regulated learning. British Journal of Educational Psychology, 75(3), 351–372. doi:10.1348/000709905X25030

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. Learning and Instruction, 12(2002), 107–119. Retrieved from papers3://publication/uuid/771F57D9-6E7D-4E98-9404-6DB8578C8EF3

Moazami, F., Bahrampour, E., Azar, M. R., Jahedi, F., & Moattari, M. (2014). Comparing two methods of education (virtual versus traditional) on learning of Iranian dental students: a post-test only design study. BMC Medical Education, 14(1), 45. doi:10.1186/1472-6920-14-45

Morgulis, Y., Kumar, R. K., Lindeman, R., & Velan, G. M. (2012). Impact on learning of an e-learning module on leukaemia: a randomised controlled trial. BMC Medical Education, 12(1), 36. doi:10.1186/1472-6920-12-36

Narjaikaew, P., Emarat, N., Arayathanitkul, K., & Cowie, B. (2010). Magnetism Teaching Sequences Based On An Inductive Approach For First-Year Thai University Science Students. International Journal of Science and Mathematics Education, 8, 891–910. doi:10.1017/CBO9781107415324.004

Nash, G., Crimmins, G., & Oprescu, F. (2016). If first-year students are afraid of public speaking assessments what can teachers do to alleviate such anxiety? Assessment & Evaluation in Higher Education, 41(4), 586–600.

Nausheen, M., & Richardson, P. W. (2013). The relationships between the motivational beliefs, course experiences and achievement among postgraduate students in Pakistan. HIGHER EDUCATION RESEARCH & DEVELOPMENT, 32(4), 603–616. doi:10.1080/07294360.2012.709485

Nevid, J. S., Pastva, a., & McClelland, N. (2012). Writing-to-Learn Assignments in Introductory Psychology: Is There a Learning Benefit? Teaching of Psychology, 39(4), 272–275. doi:10.1177/0098628312456622

Nicholas, M., van Bergen, P., & Richards, D. (2015). Enhancing learning in a virtual world using highly elaborative reminiscing as a reflective tool. Learning and Instruction, 36(2015), 66–75. doi:10.1016/j.learninstruc.2014.12.002

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. Learning and Instruction, 19(3), 259–271. doi:10.1016/j.learninstruc.2008.05.002

Özgüngör, S. (2009). The relationships between students' evalautions of teaching behaviors and self efficacy Beliefs. Procedia - Social and Behavioral Sciences, 1(1), 2687–2691. doi:10.1016/j.sbspro.2009.01.475

Peets, A. D., Coderre, S., Wright, B., Jenkins, D., Burak, K., Leskosky, S., & McLaughlin, K. (2009). Involvement in teaching improves learning in medical students: a randomized cross-over study. BMC Medical Education, 9, 55. doi:10.1186/1472-6920-9-55

Peets, A. D., Cooke, L., Wright, B., Coderre, S., & McLaughlin, K. (2010). A prospective randomized trial of content expertise versus process expertise in small group teaching. BMC Medical Education, 10(1), 70. doi:10.1186/1472-6920-10-70

Peine, A., Kabino, K., & Spreckelsen, C. (2016). Self-directed learning can outperform direct instruction in the course of a modern German medical curriculum - results of a mixed methods trial. BMC Medical Education, 16(1), 158. doi:10.1186/s12909-016-0679-0

Qureshi, S., Vishnumolakala, V. R., Southam, D. C., & Treagust, D. F. (2016). Inquiry-Based Chemistry Education in a High-Context Culture: a Qatari Case Study. International Journal of Science and Mathematics Education, (May). doi:10.1007/s10763-016-9735-9

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. Learning and Instruction, 12(5), 529–556. doi:10.1016/S0959-4752(01)00030-5

Richmond, A. S., Murphy, B. C., Curl, L. S., & Broussard, K. A. (2014). The Effect of Immersion Scheduling on Academic Performance and Students' Ratings of Instructors. Teaching of Psychology, 42(1), 26–33. doi:10.1177/0098628314562675

Roberts, J. C., & Roberts, J. C. (2015). Evaluating the Effectiveness of Lecture Capture : Lessons Learned from an Undergraduate Political Research Class Evaluating the Effectiveness of Lecture Capture : Lessons Learned from an Undergraduate Political Research Class. Journal of Political Science Education, 2169(October), 45–60. doi:10.1080/15512169.2014.985104

Rondon, S., Sassi, F. C., & Furquim de Andrade, C. R. (2013). Computer game-based and traditional learning method: a comparison regarding students' knowledge retention. BMC Medical Education, 13(1), 30. doi:10.1186/1472-6920-13-30

Saville, B. K., Bureau, A., Eckenrode, C., Fullerton, A., Herbert, R., Maley, M., … Zombakis, J. (2014). Interteaching and lecture: A comparison of long-term recognition memory. Teaching of Psychology, 41(4), 325–329. doi:10.1177/0098628314549704

Schönwetter, D. J., Clifton, R. A., Perry, R. P., & Schonwetter, D. J. (2002). Content familiarity: Differential impact of effective teaching on student achievement outcomes. Research in Higher Education, 43(6), 625–655.

Shaffer, D. M., & Collura, M. J. (2009). Evaluating the Effectiveness of a Personal Response System in the Classroom. Teaching of Psychology, 36(4), 273–277. doi:10.1080/00986280903175749

Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: the case for explicitly developing students "assessment literacy." Assessment & Evaluation in Higher Education, 38(1), 44–60. doi:10.1080/02602938.2011.598636

Tan, N. C., Kandiah, N., Chan, Y., Umapathi, T., Lee, S., & Tan, K. (2011). A controlled study of team-based learning for undergraduate clinical neurology education. BMC Medical Education, 11(1), 91. doi:10.1186/1472-6920-11-91

Tomcho, T. J., & Foels, R. (2012). Meta-Analysis of Group Learning Activities: Empirically Based Teaching Recommendations. Teaching of Psychology, 39(3), 159–169. doi:10.1177/0098628312450414

Travis, L. L., Hudson, N. W., Henricks-Lepp, G. M., Street, W. S., & Weidenbenner, J. (2016). Team-Based Learning Improves Course Outcomes in Introductory Psychology. Teaching of Psychology, 43(2), 99–107. doi:10.1177/0098628316636274

Trigwell, K., & Prosser, M. (1991). Improving the Quality of Student Learning : The Influence of Learning Context and Student Approaches to Learning on Learning Outcomes. Higher Education, 22(3), 251–266.

Trigwell, K., Ashwin, P., & Millan, E. S. (2013). Evoked prior learning experience and approach to learning as predictors of academic achievement. British Journal of Educational Psychology, 83(3), 363–378. doi:10.1111/j.2044-8279.2012.02066.x

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? Cognitive Science: A Multidisciplinary Journal, 31(1), 3–62. doi:10.1080/03640210709336984

Vittengl, J. R., Bosley, C. Y., Brescia, S. a, Eckardt, E. a, Neidig, J. M., Shelver, K. S., & Sapenoff, L. a. (2004). Why are some undergraduates more (and others less) interested in psychological research? Teaching of Psychology, 31(2), 91–97. doi:10.1207/s15328023top3102

Wang, J.-S., Pascarella, E. T., Nelson Laird, T. F., & Ribera, A. K. (2014). How clear and organized classroom instruction and deep approaches to learning affect growth in critical thinking and need for cognition. Studies in Higher Education, 5079(August 2016), 1–22. doi:10.1080/03075079.2014.914911

Wilson, J. H., & Ryan, R. G. (2013). Professor–Student Rapport Scale: Six Items Predict Student Outcomes. Teaching of Psychology, 40(2), 130–133. doi:10.1177/0098628312475033

Wilson, K., Lizzio, A., & Ramsden, P. (1997). The Development, Validation and Application of the Course Experience Questionnaire. Studies in Higher Education, 22(1), 33–53.

Yang, L.-H., Jiang, L.-Y., Xu, B., Liu, S.-Q., Liang, Y.-R., Ye, J.-H., & Tao, E.-X. (2014). Evaluating team-based, lecture-based, and hybrid learning methods for neurology clerkship in China: a method-comparison study. BMC Medical Education, 14(1), 98. doi:10.1186/1472-6920-14-98

Yoder, J. D., & Hochevar, C. M. (2005). Encouraging active learning can improve students' performance on examinations. Teach Psychol, 32(2), 91–95.

Zafra-Gómez, J. L., Román-Martínez, I., & Gómez-Miranda, M. E. (2015). Measuring the impact of inquiry-based learning on outcomes and student satisfaction. Assessment & Evaluation in Higher Education, 40(8), 1050–1069. doi:10.1080/02602938.2014.963836