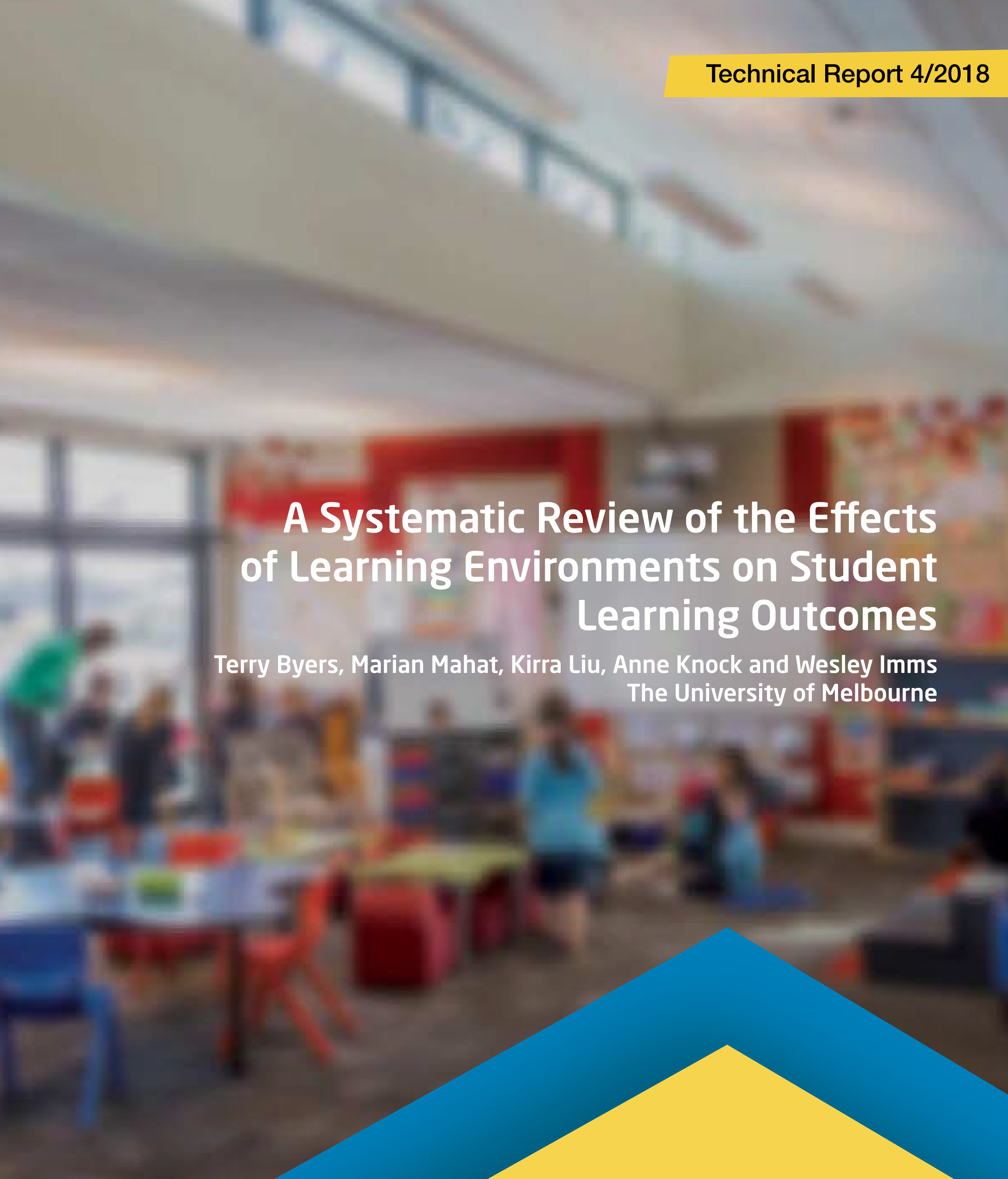

Technical Report 4/2018

A Systematic Review of the Effects of Learning Environments on Student Learning Outcomes

Terry Byers, Marian Mahat, Kirra Liu, Anne Knock and Wesley Imms
The University of Melbourne

Australian Government
Australian Research Council
iletc
LEaRN
learnetwork.edu.au
THE UNIVERSITY OF MELBOURNE

A Systematic Review of the Effects of Learning Environments on Student Learning Outcomes

Byers, T., Mahat, M., Liu, K. & Knock, A. & Imms, W. (2018). *Systematic Review of the Effects of Learning Environments on Student Learning Outcomes*. Melbourne: University of Melbourne, LEaRN. Retrieved from: http://www.iletc.com.au/publications/reports

ISBN: 978-0-7340 5502-6

ARC Linkage project (2016-2019)

This research is supported by Australian Research Council's Linkage Projects funding scheme (project LP150100022). The views expressed herein are those of the authors and are not necessarily those of the Australian Research Council

Acknowledgements

We would like to acknowledge the literature contributions of graduate researchers in the ILETC project and for their participation at various stages of the systematic review process. We would also like to thank the ILETC Chief Investigators for their input at various stages of phase 1 of the project.

Design and layout: Lachlan Stewart.

Cover images: (front) Pegasus Primary School, Jasmax Architecture, Stephen Goodenough photography; (rear) Our Lady of Sion College, Law Architects, Adam Thwaites photography.

Lead Chief Investigator
Associate Professor Wesley Imms

Chief Investigators
Professor David Clarke
Dr Ben Cleveland
Associate Professor Kenn Fisher
Professor Lisa Grocott
Professor John Hattie
Professor Thomas Kvan
Associate Professor Clare Newton

Project Manager
Joann Cattlin

Lead Research Fellow/Research Manager
Dr Marian Mahat

Research Fellows
Chris Bradbeer
Dr Terry Byers

Research Assistants
Kirra Liu
Roz Mountain
Lachlan Stewart

Graduate Researchers
Raechel French
Anne Knock
Victoria Leighton
Daniel Murphy
Mark Osborne
Dion Tuckwell
Ethel Villafranca
Pamela Yang
Fiona Young

TECHNICAL REPORT

A Systematic Review of the Effects of Learning Environments on Student Learning Outcomes

Assumption College, Y2 Architecture, Peter Clarke photography.

Overview

Aim

The systematic review identified evidence that different learning environments (blended, innovative learning environment (ILE), open-plan and traditional) have an impact on student learning outcomes. There are significant methodological questions around the availability and viability of empirical evidence. This systematic review investigated how researchers measure changes in academic outcomes attributed to the intervention of changes to the primary and secondary schooling learning environments.

Method

A search of twelve databases, which integrated fields of education or design, identified those studies that addressed student learning outcomes in a range of environments in both primary and secondary educational settings. Quantitative data was extracted using a customised form, with the application of various processes to assess bias, reliability and validity to document changes in discrete measure/s of academic or learning outcomes.

Results

Of the 5,521 articles retrieved, 21 were included in this review. The studies ranged from single-site comparative studies through to quasi-experimental randomised designs at multiple sites. Samples ranged from 17 to 22,679 students from primary and secondary schools. The review revealed that assessment regimes that favoured the prevailing view of academic progress in the domains of literacy and numeracy were most common. Importantly, the review identified few robust and valid instruments that assessed the impact of different spatial layouts on student learning in the 21st Century learning domains of creativity, critical thinking, communication, collaboration and problem-solving.

Interpretation

The review presented a small number of studies with adequate quality, sampling and statistical process to isolate and then evaluate the impact of different learning environment types. These studies presented evidence of a positive correlation between learning environments, and improvements in, student academic achievement. At the same time, the review highlighted the need for further longitudinal evaluation of how different learning environments impact a broader spectrum of student academic outcomes.

Contents

Anglican Church Grammar School Centenary Library, Brand + Slater Architects. Christopher Frederick Jones photography.

Introduction

Rationale

School learning environments are a matter of global policy and systemic government investment (Dumont & Istance, 2010). The strategic reconsideration of school learning spaces is a response to demographic, economic and technological changes that have altered the perceptions of what constitutes effective teaching and learning (see MCEETYA, 2008; New Zealand Ministry of Education, 2011; New Zealand Ministry of Education, 2014; OECD, 2013). The narrative of '21st Century Learning' (creativity, critical thinking, communication, collaboration and problem-solving) has prompted some to question the efficacy of existing classroom models and to put forward blended, open and, more recently, innovative learning environments (ILEs) (See examples of Alterator & Deed, 2013; Benade, 2017; Dovey & Fisher, 2014; Dumont & Istance; Imms, Cleveland, & Fisher, 2016). Debates around the form and function of what constitutes an effective learning environment have led some Organisation for Economic Cooperation and Development (OECD) countries, such as Australia and New Zealand, to invest significant public funding in new school buildings. In Australia alone, more than AUS$16 billion was approved for investment in school building projects from 2009 (Wall, 2009).

Despite the current interest and systemic investment in school learning environments, there is a lack of empirical data to adequately evaluate how existing and alternative learning environments (blended, ILEs and open) impact teaching and learning (Blackmore, Bateman, O'Mara, & Loughlin, 2011; Brooks, 2011; Gislason, 2010). Brooks is critical of the overt theorising around these new spaces, with a "dearth of systematic, empirical research being conducted" on their impact on teaching and learning (p. 719). For Painter, Fournier, Grape, Grummon, Morelli, Whitmer, & Cevetello (2013), this lack of evidence stems from the fact that there are very few methodologies and metrics able to isolate and then assess how different learning spaces affect both teachers and students.

There remains little understanding if, or to what extent, different school learning environments affect student academic or learning outcomes (Blackmore et al.).

This systematic review draws on studies from the integrated fields of education and design. It focuses on identifying quantitative studies with valid methodologies that isolate the variable of different learning environment type/s (blended, ILEs, open-plan and traditional) and analyse their impact on reliable measures of student academic achievement. The ensuing analysis tests the suggestion that there is currently a lack of substantive, empirical data around the claims that different learning environment types correlate, whether positively or negatively, with student academic or learning outcomes.

Objectives

In the context of learning environment type/s (blended, ILEs, open-plan and traditional), the objectives of this review are:

1. To identify studies that investigate the impact of learning environment types on student learning outcomes;
2. To identify measures of student learning outcomes or academic achievement;
3. To determine the content of the identified student learning outcome measures (those with published evidence of reliability and validity).

The results of this study will provide researchers and practitioners with a better understanding of how currently available student learning outcome measures are quantified in the context of learning environments.

Research Questions

1. *What evidence exists that different learning environments have an impact on student learning outcomes?*
2. *What student learning outcome measurement tools have been designed and used for measuring student outcomes in different learning environment types?*
3. *What elements of student learning are quantified by the identified measurement tools?*

Methods

Review methodology

The review adopted the principles and techniques of systematic reviews, which involved sifting abstracts, scrutinising full papers and abstracting data. One researcher performed the initial search and subsequent data extraction. Two members of the team checked each title and abstract to decide whether the full paper should be read. The lead team member was consulted if a difference of opinion arose. Similarly, each full paper was read by at least two members of the team and agreement sought from the lead team member for any variations of opinions. Two other members of the team checked 10% of the abstraction records. The complexity of the underlying construct of different learning environments and student learning outcomes and the breadth of studies retrieved in the search necessitated a full-team discussion to determine the final list of included articles. The potential for selection bias was addressed by the disciplined process followed by reviewers, sometimes bringing into the discussion their or another team member's specific understanding of the construct..

Search strategy

The database search for this systematic review was performed in January 2017 using EBSCOhost databases (Academic Search Complete, Avery Index to Architectural Periodicals, Education Research Complete, Educational Administration Abstracts, ERIC), Proquest databases (Education Database, Art and Humanities Database, Humanities Index, PAIS), OVID, Informit, Scopus and Web of Science. These databases integrate information from the fields of education and design and include articles addressing student learning outcomes in a range of environments. A study protocol was not registered. Operational definitions were determined for each aspect of our search, which related directly to the research question and selection criteria (Table 1). The search terms were developed using related literature and chosen by team consensus based on their theoretical and practical significance. Search terms addressed the concepts of student learning outcomes and different learning environment type/s, and the use of intervention-based study designs. Where available, exploded search terms were used, as well as associated terminology in the title, abstract, and, where appropriate, the keywords of the articles. Boolean operators helped narrow the search to relevant research fields.

Selection criteria

An ILE is defined by the OECD (2013, p. 11) as "an organic, holistic concept – an ecosystem that includes the activity and the outcomes of the learning". The concept embraces the learning taking place as well as the setting. The OECD describes ILEs as multi-modal, technology-infused and flexible learning spaces that are responsive to evolving educational practices (OECD, 2015). The selection criteria for this review (Table 2) aimed to document how researchers perceived student learning outcomes and therefore focused on the definitions they provided. To ensure the inclusion of a comprehensive breadth of articles, the application of the selection criteria did not use operational definitions of student learning outcomes. The population of interest in this review was both primary and secondary school students. Therefore, studies were limited to those which involved students in these schools. To ensure that relevant information was not missed, studies were included that considered known student learning outcome measures, even if student learning outcomes were not the primary study objective. In the current review, the aim was to assess quantitative changes in student learning outcomes before and after assigned intervention(s); therefore, articles were included on the basis of discrete measure/s of academic or learning outcomes.

Table 1: Study search terminology

Topic	Search terms	Exploded search terms (abstract/title)
Population	Elementary and Secondary School students	“Elementary school student*” “Elementary student*” “Primary school student*” “Secondary school student*” “Secondary student*” “High school student*” “Junior high school student*” “Junior high student*” “Middle school student*” Pupil* Schoolchild* Schoolboy* Schoolgirl*
Student Learning Outcome	Student learning outcomes	“Student learning outcome*”
	Academic achievement	“Academic achievement*”
	Academic outcomes	“Academic outcome*”
	Academic success	“Academic success*”
	Educational achievement	“Educational achievement*”
	Educational outcomes	“Educational outcome*”
	Grade point average	“Grade point average*” GPA*
	Student outcomes	“Student outcome*”
Innovative Learning Environments	Innovative learning space Modern learning space Contemporary learning space Physical learning space 21st century learning space	“Learning space*”
	Innovative learning environment Modern learning environment Contemporary learning environment Physical learning environment 21st century learning environment	“Learning environment*”
Contemporary learning space	Innovative learning environment Modern learning environment Contemporary learning environment Physical learning environment 21st century learning environment	“Learning environment*”
	Physical environment	"Physical environment*"
	Physical space	"Physical space*"
	School environment	"School environment*"
	School space	"School space*"
	Classroom environment	"Classroom environment*"

Table 2: Slection criteria

Search	Include	Exclude
Population	Primary and secondary students	Kindergarten and post-secondary students
Outcomes of selected studies	Quantitative	Qualitative
Design	Intervention-based studies including randomised controlled trial, quasi-randomised controlled trial, single-group pre- and post-test studies, and single case experimental designs.	Other systematic reviews or literature reviews
Publication type	Articles published as full texts in peer review journals	Articles or abstracts not published in peer review journals Articles or abstracts published in languages other than English Grey literature (conference proceedings theses, books and other grey literature) Systematic reviews (although reference lists were used to ensure all relevant publications were located) Generalised discussion papers of participation measures that did not present new evidence from a scientific study Qualitative studies as the focus was on quantitative outcome measures
Time	1960 - 2016	Prior to 1960

Data collection and assessment of quality

Covidence was the primary screening and data extraction tool used for the systematic review.

The evaluation of sampling bias, reliability and validity ascertained the quality of selected studies. The *Cochrane Collaboration* tool assessed the risk of selection, detection, attrition and reporting bias (Higgins et al., 2011). Even though the tool was developed specifically for assessing randomised control trials, it presented a viable means to evaluate a wider variety of methodological designs of the selected studies. A summary of the relevant measures utilised by the selected studies was summarised to outline key characteristics, differences and similarities across the final study selection. The work of Campbell and Stanley (1963) informed the assessment of the internal validity (history, instrumentation, maturation and selection). While, assessments of internal consistency and reliability were made on the relevant aspects of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (Terwee et al., 2012).

Synthesis of studies

A two-part narrative approached analysed the results of the selected studies. First, the research questions of this review framed the analysis of findings of the selected studies. Second, a synthesis of the collective results addressed potential gaps and issues and established a frame for future meta-analysis.

Systematic Review Results

The initial database search revealed 5,521 articles after applying filters based on the selection criteria: primary and secondary students in quantitative intervention-based and single-case experimental studies, published as full texts in peer review journals between 1990 and 2016. Figure 1 displays the number of references yielded during the initial database search and subsequent stages of the review. Following this identification, the number of references was reduced to 4,481 after the removal of duplicates, articles in languages other than English, non-peer reviewed journal articles and articles published before 1990. Appraisal of titles and abstracts excluded 4,409 articles, with 72 articles undergoing full-text review. Only 21 of these made the final analysis; 51 did not have student academic or learning outcomes as a dependent variable or statistically analysed changes in measures of achievement..

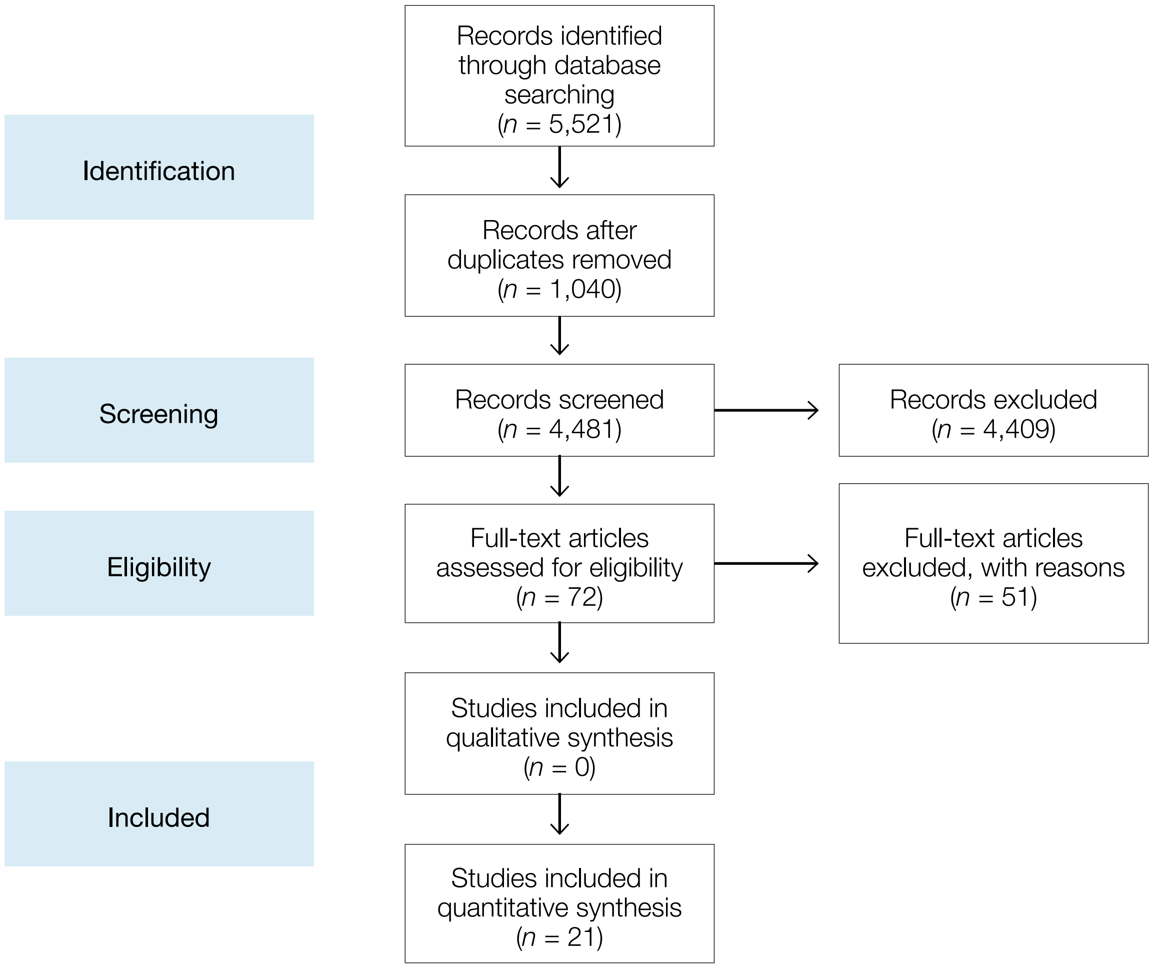

Figure 1. PRISMA flow diagram of the articles yielded during the systematic review process (including removal of duplicates and references that did not align with selection criteria). The demographic descriptors of participants in each of the studies, the student outcomes measured and study characteristics are shown in Table 3.

Table 3: Demographic descriptors, student outcomes measured and study characteristics.

Author, Year, Title First	Sample Characteristics				Student Outcome Measures		Instrument Use	Theory	Focus of study	LE	Effects on SO	Effect size
	n	Age (M, SD, R[a])	Sex (M[b]:F[c])	Level Primary (P) Secondary (S)								
Bottge, B., (2006). Situating math instruction in rich problem-solving contexts: Effects on adolescents with challenging behaviors	Total = 17	R = 15-18	10:7	S[d]	Maths achievement on various tests.		Multimedia aligned tests: Fraction of the Cost Challenge (FCC), Kim's Komet Challenge (KKC) Vs. Traditional: Fractions computation test, Standardized tests.	*ND*[f]	To examine the effects of enhanced anchored instruction (EAI) on the maths achievement (particularly fraction knowledge) of students of low achievement due to challenging behaviours.	Generative maths LE[f]	Students performed better on both curriculum aligned tests: FCC & KCC which were multimedia and hands-on, but showed no improvement in the fractions computation test and standardised tests.	FCC 0.75 KKC 0.78
Chandra, V., (2008). The methodological nettle: ICT and student achievement	Total = 233	R = 15-16	ND (7:7)	S	Achievement on test scores		Test scores	*ND*	To assess the achievement of two groups of students over two years: one traditional, and one blended, to determine whether ICT has an impact on achievement.	Trad.[h] LE Vs blended or e-learning LE	The blended LE improved student achievement in comparison to the trad. However, the comfort with blended LE differed and was independent of test-score.	*ND*
Cicek, F .G., (2016). Laboratory Control System's effects on student achievement and attitudes.	Total = 66 Experiment = 33 Control: 33	R = 15-16		S	Academic achievement of students in Web Design		Scale of Attitudes towards Learning and Teaching Process and the Achievement Tests (pre- and post-tests)	*ND*	To determine the academic achievement of students in Web Design in an LCS and Trad. LE	Laboratory control system (LCS) vs Trad. LE	Post-tests of academic achievement, retention and positive attitude after doing Web design were higher in the LCS LE. The Trad. LE also improved achievement.	*ND*
Forman, S. G., (1978). Creativity and achievement of second graders in open and traditional classrooms	Total = 129 Open = 63 Trad = 66	R = 6-7	(66:63) Open (32:31) Trad. (34:32)	P[g]	IQ, creativity, achievement		Primary Mental Abilities Test (PMAT), Iowa Tests of Basic Skills	*ND*	To compare the level of creativity and achievement of students in open LE vs Trad. LE	Open vs closed	Students in traditional classrooms scored higher in fluency (in one school system), vocabulary, and reading and mathematics achievement. No significant differences for uniqueness.	*ND*
Fößl, T., (2016). A field study of a video supported seamless-learning-setting with elementary learners	Total = 85 Experiment = 24 Control = 23 FC1 = 25 FC2 = 13	M = 10.6 SD = .31	(77:8)	P	Learning performance		Learning performance (pre- and post-test), Video views, Working progress, survey, student & teacher interview, observations.	*ND*	To compare the learning performance of students in a seamless real-world learning setting (using open learning and video) and a Trad. LE	Open-learning approach with video-supported seamless-learning-setting	The students in the experimental/seamless LE setting performed better on the post-test than the students in a Trad. LE.	0.31
McRobbie, C. J., (1993). Associations between student outcomes and psychosocial science environment.	Total = 1,594 Inquiry and LE =591 Attitude and LE = 596	R = 12-18	*ND*	S	Student outcomes: student and class mean scores		The Science Laboratory Environment Inventory	*ND*	To investigate the correlation between the psychosocial environment and their effect on student outcomes	Science LE	Student outcomes were found to be affected by psychosocial environment-independent of ability. Attitude and inquiry skill outcomes were enhanced when laboratory activities were integrated into non-laboratory classes	*ND*
Ozerbas, M. A., (2016). The Effect of the Digital Classroom on academic success and online technologies self-efficacy.	Total = 58 Experiment = 32 Control = 26	R = 12-13	(31:27)	S	Academic success		Pre and post-test of the Academic success test and Online Technologies Self-Efficacy Scale	*ND*	To determine the effect of digital classrooms on the academic success of students and technology self-efficacy.	Digital LE	The experiment group in the digital classroom has significantly higher academic success than the controls (no access to digital technology). There was no effect on online technology self-efficacy.	*ND*

Author, Year, Title First	Sample Characteristics				Student Outcome Measures		Instrument Use	Theory	Focus of study	LE	Effects on SO	Effect size
	n	Age (M, SD, R[a])	Sex (M[b]:F[c])	Level Primary (P) Secondary (S)								
Reiss, S., (1975). Persistence, achievement, and open-space environments	Total = 182 (30 per six schools approx.) Open = 85 Closed = 88	R = 7-8	(15:15) per school	P	Persistence, achievement		Behavioural post-test for persistence on a difficult task, the Stanford Preschool Internal-External (I-IC) Scale, and the California Achievement Test (CAT)	*ND*	To examine the difference in persistence and achievement in open-space LE's vs Trad. LE's.	Open vs Trad. LE	Open LE's promoted higher persistence on difficult tasks than the traditional LE across all genders. Persistence and achievement were more positively correlated in open LE than Trad. LE for boys only.	*ND*
Shamaki, T. A., (2015). Influence of Learning Environment on Students' Academic Achievement in Mathematics: A Case Study of Some Selected Secondary Schools in Yobe State – Nigeria	Total = 337	ND	ND	S	Academic achievement		structural questionnaire as well as an achievement test	*ND*	To investigate the influence of the LE on mathematics students' academic achievement.	Ideal vs dull LE	Student performance was negatively correlated with dull painting, over-crowdedness and poor lighting and positively correlated with adequate seating/room. Adequate ventilation had no effect	*ND*
Solomon, D., (1976). Individual Characteristics and Children's Performance in "Open" and "Traditional" Classroom Settings.	Total = 92	R = 9-10	(56:36)	P	Academic achievement, creativity, inquiry skills, attitudes, behaviour.		Classroom observations, questionnaires, Iowa Tests of Basic Skills	*ND*	To examine children's characteristics and performance in open and traditional settings.	Open vs Trad. LE	Children in open classes scored significantly higher in creativity, democratic/co-operative behaviour and involvement in class activities but significantly lower on the achievement test, undisciplined activity level and social involvement.	*ND*
Tanner, C. T., (2008). Explaining Relationships Among Student Outcomes moreover, the School's Physical Environment.	Total = 1,916 (24 schools)	R = 8-9	*ND*	P	Academic achievement: Iowa Test of Basic Skills score (ITBS)		Iowa Tests of Basic Skills	*ND*	To investigate the effects of four sets of design patterns in the physical school environment: movement and circulation (MC), large group meeting places (LGMP), daylight and views (DLV), and instructional neighbourhoods (IN) on achievement	Four different design patterns: ML, LGMP, DLV, IN.	As school design score increased, so did achievement on the ITBS in each of 24 schools.	MC 0.069 LGMP 0.018 DLV 0.025 IN 0.031
Uline, C., (2008). The walls speak the interplay of quality facilities, school climate, and student achievement.	Total = 1,134 (80 schools	ND	*ND*	M	Academic achievement		Teacher surveys, student SES and achievement data.	*ND*	To examine the mediating role of school climate on the facility quality level and student achievement and the interplay of these factors.	LE with variable facilities quality level	School facility quality was mediated by school climate variables: academic press, teacher professionalism and community engagement, and was found to affect student achievement. Inadequate facilities meant a focus on academics were lowered & teachers were less enthusiastic.	*ND*
Kazu, I. Y., (2014). The effect of blended learning environment model on high school students' academic achievement.	Total = 54 Experiment = 27 Control =37	R = 17-18	(37:17)	S	Academic performance		Pre and post-knowledge test, prior achievement grades	*ND*	To compare students' academic achievement in a Blended LE and a Trad. LE	Blended LE vs Trad. LE	Students learning in the blended LE had better academic outcomes than those in the Trad. LE	*ND*

Author, Year, Title First	Sample Characteristics				Student Outcome Measures		Instrument Use	Theory	Focus of study	LE	Effects on SO	Effect size
	n	Age (M, SD, R[a])	Sex (M[b]:F[c])	Level Primary (P) Secondary (S)								
Barrett, P., (2015). The impact of classroom design on pupils' learning: Final results of a holistic, multi-level analysis	Total = 3,766 Year 1 = 447 Year 2 = 606 Year 3 = 744 Year 4 = 656 Year 5 = 708 Year 6 = 606	R = 5-11	(1,883:1,883)	P	Reading, Writing and Mathematics progress points added to create an Overall Progress score		National Curriculum Key Stages 1 and 2 tests	*ND*	Assessments of the design parameters of Naturalness (Light, Temperature, Air Quality), Individualisation (Ownership, Flexibility) and Stimulation (Complexity and Colour)G were made of 153 classrooms in 27 schools in order to identify the impact of the physical classroom features on the academic progress of the 3766 pupils who occupied each of those specific spaces.	Existing Classrooms	Correlations of Overall Progress for each pupil against environmental measures showed all ten parameters were positively correlated with progress. Multilevel (Two level) modelling portioned between pupils (highest) and classroom levels. School level accounted for little (3%) of variance. Naturalness parameters (light, temperature and air quality) highest effect of 28%. Variation of the most effective classroom (Overall progress of 16.05 NC points) minus least effective (8.12 NC points), results in 7.93 NC points for variation. The impact of the classroom environmental factors, therefore, models at 7.93/50 (50 points is the total) explains 16% of the variation in pupils' academic progress achieved.	*ND*
Barrett, P., (2017). The holistic impact of classroom spaces on learning in specific subjects	Total = 3,766 (same as 2015 study) Blackpool = 715 Hampshire = 1,535 Ealing = 1,480	R = 5-11	(1,883:1,883)	P	Reading, Writing and Mathematics progress points added to create an Overall Progress score		National Curriculum Key Stages 1 and 2 tests	*ND*	Assessments of the design parameters of Naturalness (Light, Temperature, Air Quality), Individualisation (Ownership, Flexibility) and Stimulation (Complexity and Colour) were made of 153 classrooms in 27 schools in order to identify the impact of the physical classroom features on the academic progress of the 3766 pupils who occupied each of those specific spaces.	Existing Classrooms	1. See above 2. % Improvement due to classroom parameters was 9.3% (Reading), 8.4% (Writing) and 11.7% (Maths) 3. For each of the different subject models, the aspects of the classroom environment taken together explained approximately 10% of the variability of pupil performance.	*ND*
Byers, T., (2014). Making the case for space: The effect of learning spaces on teaching and learning	Total = 386 *n* = 164 (6 classes in intervention) *n* = 222 (control group who remained in traditional with only learning outcomes compared	R = 11-14	(164:0)	S	English and Mathematics academic achievement (A+ to E-), Verbal and Non-Verbal Reasoning standardised data (as a proxy of cognitive ability) Student attitudinal		English and Mathematics school-based assessment Academic Assessment Services Cognitive Ability Tests Linking Pedagogy, Technology and Space repeated measures survey	*ND*	Examined the impact of learning spaces on teachers' pedagogy, student engagement and student learning outcomes in a technology-rich school setting. Its quasi-experimental design allowed examination of differences in these variables between two settings – 'traditional' classrooms, and 'new generation learning spaces.' (NGLS).	Existing Trad. LE (control) and Retrofitted NGLS-model ILE affordances	A two-tailed, paired t-test with an alpha level of 0.05 compared participating students' assessments in these subjects taken during the time they occupied the traditional and the NGLS classrooms. The null hypothesis was rejected with nine out of the twelve results showing a statistically significant improvement in student learning outcomes. These statistically significant improvements were justified by Cohen's d effect size, ranging from the upper end of the small to high effects.	Within *d* =.40 in English Within *d* =.41 in Maths Overall *d* =.40

Author, Year, Title First	Sample Characteristics				Student Outcome Measures		Instrument Use	Theory	Focus of study	LE	Effects on SO	Effect size
	n	Age (M, SD, R[a])	Sex (M[b]:F[c])	Level Primary (P) Secondary (S)								
Chang, C. Y., (2006). Preferred - Actual learning environment “spaces” and earth science outcomes in Taiwan	Total = 155	Mean of 16 years	(74:81)	S	Earth Science Learning Outcomes Inventory -Students’ perceptions on preferred/actual learning environment and the students’ learning achievement and attitude		Earth science classroom learning environment instrument (ESCLE) and Earth Science Learning Outcomes Inventory (ESLOI)	*ND*	Determine whether preferred–actual space accounts for outcome variance beyond that explained by actual learning environment.	traditional (teacher-centred) learning environment (TLE) and mixed (both teacher- and student-centred learning environment (MLE)	Results indicated that although preferred–actual space is not related to achievement, but is both statistically and practically associated with attitudes toward the subject when actual learning environment is controlled. regression analysis revealed that the pre-test scores were the only significant predictors in explaining students’ learning outcomes	*ND*
Chang, C. Y., (2011). Science Learning Outcomes in Alignment with Learning Environment Preferences.	Total = 155	Mean of 16 years	*ND*	S	Earth Science Learning Outcomes Inventory -Students’ perceptions on preferred/actual learning environment and the students’ learning achievement and attitude		Earth science classroom learning environment instrument (ESCLE) and Earth Science Learning Outcomes Inventory (ESLOI)	*ND*	Aimed to investigate students’ learning environment preferences and to compare the relative effectiveness of instructional approaches to students’ learning outcomes in 10th-grade earth science classes.	Student–Teacher-Balanced Instructional Model, STBIM) and teacher-centred group (the Teacher-Centred Instructional Model, TCIM).	Students preferred a classroom environment where student-centred and teacher-centred instructional approaches coexisted (STBIM) over a teacher-centred (TCIM) learning environment. It was also revealed that the STBIM students’ achievement in and attitude toward earth science were enhanced when the learning environment was congruent with their learning environment preference,	*ND*
Gilavand, A., (2016). Investigating the impact of schools’ open space on learning and educational achievement of elementary students	Total = 210	R = 7-11	*ND*	P	Student Attitudinal Data Observational Data		Construction Observational Checklist Academic Achievement Motivation Questionnaire	*ND*	This study investigated the impact of schools’ open space on learning and educational achievement of elementary students in Ahvaz, Southwest of Iran.	Open spaces – exterior play spaces	schools’ open space has a significant impact on learning and academic achievement (ascertained through a questionnaire) of elementary school students in Ahvaz- Iran	*ND*
Tanner, C. K. (2000). The influence of school architecture on academic achievement	Total = 22, 679	R = 10 - 11	*ND*	P	Reading and Mathematics scores		Iowa Test of Basic Skills (ITBS) University of Georgia’s School of Design and Planning Laboratory (SD & PL) scale	*ND*	Focused on the neglected aspect of the physical environment of the classroom impact. Its purpose was to determine how school architectural design factors might influence student achievement scores in elementary schools	Used SD & PL scale to assess existing classroom spaces	A total of 7 design factors were found to correlate to student learning outcomes (compatibility with context, clearly defined pathways, positive outdoor spaces, computers for teachers, positive overall impression had the greatest impact on ITBS scores	*ND*

Note. [a] Age range; [b] Number of males; [c] Number of females; [d] Secondary years of schooling; [e] Learning environment; [f] Not disclosed; [g] Primary years of schooling; [h] Traditional

Participants

The number of student participants in each of the studies ranged from 17 to 22,679 students (average of 1,665). The age of students in the final studies ranged from 5-18yrs. Equitable distribution of boys and girls occurred in the respective samples, with the only exception being the Byers, Imms, and Hartnell-Young (2014) study (with an all-boys school as the site). In many instances, authors either undertook a randomised selection process or ensured the sample reflected key student demographic characteristics (i.e. ethnicity, socio-economic status and location) with equitable distribution between control and intervention groups. A high proportion of the selected studies engaged in pre-testing of student academic performance to show that they were statistically similar. Consequently, these sampling measures moderated the incidence of selection bias that would distort the statistical analysis of the between-group comparisons.

Outcomes and Measures

The student outcomes identified in this review were measured using computations of observed participation behaviours, a variety of standardised tests, general achievement tests, prior achievement data and participant surveys. The range of measures used to determine student outcomes is displayed in Table 4.

The assessment of student learning outcomes covered a range of assessment devices and types across the 21 studies. Seven studies utilised school-based assessment to determine the impact of different learning environment types on student academic achievement. These school-based assessments were in English, Mathematics and Science (see Byers et al., 2014; Chandra & Lloyd, 2008; Cicek & Taspinar, 2016; Kazu & Demirkol, 2014; Shamaki, 2015). The majority of studies utilised an amalgam of standardised test measures. Most prominently used was the Iowa Test of Basic Skills (ITBS), followed by the California Achievement Test (CAT) and the English National Curriculum Key Stages 1 and 2 tests. The external, standardised nature of these items presented a reliable and valid means to facilitate generalisable analysis across multiple schools (see Barrett, Davies, Zhang, & Barrett, 2015, 2017; McRobbie & Fraser, 1993; Reiss & Dyhdalo, 1975; Tanner, 2000, 2008; Uline & Tschannen-Moran, 2008).

Even though this review focused on measures around student academic learning outcomes, many utilised assessment of student attitudes to learning and observations of students in various spatial layouts. Some used repeated measure surveys to elicit student attitudinal responses (see Byers et al., 2014; Chang, Hsiao, & Barufaldi, 2006; Chang, Hsiao, & Chang, 2011; Cicek & Taspinar, 2016; Fößl, Ebner, Schön, & Holzinger, 2016; Gilavand, Espidkar, & Gilavand, 2016). In others, comparative observations of traditional learning environments and ILEs (Fößl et al., 2016; Gilavand et al., 2016; Solomon & Kendall, 1976) focused on discerning the differences in pedagogies and learning experiences. Finally, a small number of studies (Barrett et al., 2015, 2017; Tanner, 2000, 2008) assessed the physical design parameters (i.e. individualisation and stimulation) and environmental factors (i.e. light, temperature and air quality) to investigate the potential impact of the physical classroom features on student academic progress.

Table 4: Student outcomes, measures and constructs of final articles.

First author, year	Measure	Student outcome	Construct
Bottge, B., (2006)	Fraction of the Cost Challenge (FCC)	Maths achievement	FCC = 36-point test with points allotted for correct working and answer
	Kim's Komet Challenge (KKC)	Maths achievement	KCC = 37-point test with points allotted via degree of difficulty
	Fractions computation test	Maths achievement	Fractions computation test
	Standardized tests- the Iowa Tests of Basic Skills (ITBS): Mathematics subtest.	Academic achievement	ITBS: Standardised achievement in Mathematics test that result in measures such as National Grade Equivalents (NGE).
Chandra, V., (2008)	Test scores	Achievement on test scores	Tests developed within the school were sat, providing performance data
Cicek, F .G., (2016)	Scale of Attitudes towards Learning and Teaching Process (SALTP)	Attitude	38-item attitude scale
	Achievement test	Academic achievement	Achievement test- consisting of 28 items
Forman, S. G., (1978)	Wallach-Kogan Test: consist of three verbal: Instances, Alternate Uses and Similarities subtest and two figural subtests: Pattern Meanings and Line Meanings	Academic achievement	Wallach-Kogan Test: Provides a Fluency score- the number of appropriate responses summed over items, and uniqueness score-total number of responses that appeared only once in the study sample for a given item.
	ITBS: Mathematics subtest	Creativity	ITBS: Standardised achievement in Mathematics test that result in measures such as National Grade Equivalents (NGE).
	Primary Mental Abilities Test (PMAT)	Achievement	PMAT
Fößl, T., (2016)	Learning performance test	Learning performance	Learning performance test: 12 single choice questions and 4 practical exercises where students had to draw something
	Students' attitude towards worked example videos (SATWEV) survey	Attitude	SATWEV-17 item, five-point Likert-type scale survey

First author, year	Measure	Student outcome	Construct
McRobbie, C. J., (1993)	The Science Laboratory Environment Inventory (SLEI): student outcome measures included four attitude measures and two inquiry skill.	Student outcomes	SLEI: Contains 35 items, with 7 assessing each of the five scales (left) which were scored on a 5-point Likert scale (almost never, seldom, sometimes, often, very often).
Ozerbas, M. A., (2016)	Academic Success Test (AST)	Academic success	AST: a multiple-choice test of 24 items and a maximum score of 100 (no points were taken away for wrong answers). Content includes 7th-grade math lessons, specific features of circles.
	Online Technologies Self-Efficacy Scale (OTSES)	Competency with the internet	OTSES: 33-item scale, with four sub-scales of Internet competencies, Synchronous interaction, Asynchronous interaction I and Asynchronous interaction II
Reiss, S., (1975)	Persistence test: A Behavioural post-test measuring persistence with puzzle-making.	Persistence	Persistent test: the average time the child worked on a puzzle (consisting of problems 1 and 2 of the Wechsler Intelligence Scale for Children (WISC) Block Design Test and Problems 9 and 10 of the Wechsler Adult Intelligence Scale (WAIS) Block Design Test).
	The Stanford Preschool Internal-External Scale (SPIES)	Expectancies about locus of control and behaviour of children in theoretically relevant situations	SPIES: is scored in the internal direction, and represent expectancies for internal control of positive events (I+) and negative events (I-) and the sum of these 2 (total I).
	The California Achievement Test (CAT)	Achievement	CAT: Standard administration in California near the end of the school year provides a percentile ranking of each student.
Shamaki, T. A., (2015)	General achievement test	Academic achievement	-

First author, year	Measure	Student outcome	Construct
Solomon, D., (1976).	CAT	Academic achievement	CAT: Standardised test administered in California near the end of the school year which provides a percentile ranking of each student.
	Virginia Standards of Learning (SOL) tests English (Reading, Research, and Literature) and Math.	Academic achievement	SOL: Standardised test that produces an achievement score.
	Virginia Test of Academic Proficiency	Academic achievement	Standardised test that produces an achievement score.
Kazu, I. Y., (2014)	Pre and post-knowledge achievement test	Academic achievement	
	Prior achievement grades	Academic achievement	Pre-test and post-test consisting of 25 and 21 items respectively, with items in compliance with the objectives of the Biology course and Bloom's taxonomy.
Barrett, P., (2015)	National Curriculum Key Stages 1 and 2 tests	Academic achievement	Reading, Writing and Mathematics progress points added to create an Overall Progress score
Barrett, P., (2017)	National Curriculum Key Stages 1 and 2 tests	Academic achievement	Reading, Writing and Mathematics progress points added to create an overall progress score
Byers, T., (2014)	English and Mathematics school-based assessment	English and Mathematics academic achievement	Achievement: A+-E-
	Student attitudinal data instrument?	Learning and engagement	A nine-item, five-point Likert scale survey measuring the effect of learning space on students' learning and engagement
	Academic Assessment Services Cognitive Ability Tests: Verbal and Non-Verbal Reasoning standardised data (as a proxy of cognitive ability)	Academic Achievement	Standardised test
Chang, C. Y., (2006)	Earth Science Learning Outcomes Inventory (ESLOI)-including the attitudes toward the earth science inventory (ATESI) and Earth Science Achievement Test (ESAT)	Students' learning achievement and attitude. ATESI: attitude towards earth science and ESAT: achievement in earth science	ESLOI: Divided into two sections with a total of 60 items. The first, is the ATESI which consists of 30 items with bipolar disagree/agree on statements on a 1–5 Likert scale, and second, the ESAT, with another 30 MCQ items.

First author, year	Measure	Student outcome	Construct
Chang, C. Y., (2011)	Earth Science Learning Outcomes Inventory (ESLOI)-including the attitudes toward the earth science inventory (ATESI) and Earth Science Achievement Test (ESAT)	Students' learning achievement and attitude. ATESI: attitude towards earth science and ESAT: achievement in earth science	ESLOI: Divided into two sections with a total of 60 items. The first, is the ATESI which consists of 30 items with bipolar disagree/agree on statements on a 1–5 Likert scale, and second, the ESAT, with another 30 MCQ items.
Gilavand, A., (2016)	Academic Achievement Motivation Questionnaire of Hermance (AAMQH)	Academic Achievement and motivation	AAMQH: 29 items based on ten (in this case 9) characteristics that distinguish those of high and low achievement motivation.
Tanner, C. K. (2000)	ITBS: Mathematics and Reading subtests	Academic achievement	ITBS: Standardised achievement in Reading and Mathematics tests that result in measures such as National Grade Equivalents (NGE).

Risk of Sampling Bias and Quality Assessment

Sampling bias

The Cochrane Collaboration tool assessed the risk of sampling bias. The tool assessed the source of bias in the domains of selection, detection, attrition and reporting on the basis of "low", "high", and "unclear" risk (Higgins et al., 2011). The studies in the final selection of this review tended to moderate sampling bias through their design and sampling process (Table 5). Random selection of moderate to large control- and intervention -groups from single or multi-sites (Barrett et al., 2015, 2017; Chandra & Lloyd, 2008; Chang et al., 2006; Chang et al., 2011; Gilavand et al., 2016; Kazu & Demirkol, 2014; Ozerbas & Erdogan, 2016; Shamaki, 2015; Tanner, 2000, 2008), decreased the incidence of selection bias. There were instances where randomisation was not employed. However, sampling bias was moderated through pre-testing of IQ, socioeconomic status or achievement scores to establish that the comparative samples were not statistically different (Byers et al., 2014; Cicek & Taspinar, 2016; Forman & McKinney, 1978; Reiss & Dyhdalo, 1975; Uline & Tschannen-Moran, 2008). Subsequent reviews of attrition, detection and reporting bias, indicated that the design, sampling strategy and method of analysis of these studies were more robust than those employing a single intervention group (Bottge, Rueda, & Skivington, 2006; Fößl et al., 2016; McRobbie & Fraser, 1993). The review highlighted the deficiency of these studies, due to their within-group design, to reliably detect significant differences at the standard of those with a comparative group design.

Quality

The assessment of internal validity and reliability established the quality of individual studies. Due to the single intervention or site designs of many studies in the final selection, the internal validity guidelines of Campbell and Stanley (1963) were applied to assess the validity in terms of history, instrumentation, maturation and selection. The remaining studies had designs that incorporated multi-sites and random selection of participants to establish generalisable

Table 5: Assessment of sampling bias of selected studies using the Cochrane Collaboration tool.

FIRST AUTHOR, YEAR	SELECTION	DETECTION	ATTRITION	REPORTING
Bottge, B., (2006)	-	-	+	+
Chandra, V., (2008)	+	+	+	+
Cicek, F .G., (2016)	+	+	+	+
Forman, S. G., (1978)	+	+	?	+
Fößl, T., (2016)	-	-	-	+
McRobbie, C. J., (1993)	-	-	?	+
Ozerbas, M. A., (2016)	+	+	-	+
Reiss, S., (1975)	+	+	-	+
Shamaki, T. A., (2015)	+	+	+	+
Solomon, D., (1976)	-	+	+	+
Tanner, C. T., (2008)	+	+	+	+
Uline, C., (2008)	+	+	+	+
Kazu, I. Y., (2014)	+	+	+	+
Barrett, P., (2015)	+	+	+	+
Barrett, P., (2017)	+	+	+	+
Byers, T., (2014)	+	+	+	+
Chang, C. Y., (2006)	+	+	+	+
Chang, C. Y., (2011)	+	+	+	+
Gilavand, A., (2016)	+	+	+	+
Tanner, C. K. (2000)	+	+	+	+

Note	+	-	?
	Low risk	High risk	Unknown risk

evidence. Regarding checks for reliability, the articles were evaluated by reported measures of internal consistency and inter-rater reliability as dictated by the COSMIN checklist (Terwee et al., 2012). The rationale for not using the COSMIN criteria to assess the validity of this selection is due to its focus on assessing large sample, randomised control trials. The checklists assessment of the quality of methodologies and measures were beyond the scope of the intervention-based design of the selected studies. The application of the COSMIN checklists four-point criterion of “excellent”, “good”, “fair” and “poor” for both reliability and validity provided an efficient means for establishing the overall assessment quality (“strong”, "moderate", "low" and "unknown") of each study's design and measures.

The quality of the selected articles ranged from low to strong, with the majority falling into the category of moderate (Table 6). The studies identified as strong in terms of quality (Barrett et al., 2015, 2017; Chang et al., 2006; Chang et al., 2011; Cicek & Taspinar, 2016; Ozerbas & Erdogan, 2016; Tanner, 2000, 2008; Uline & Tschannen-Moran, 2008) were best described as large, multi-site, randomised comparative studies that tended to utilise assessment through existing external, standardised testing instruments. Not only were their designs rigorous, but they utilised, and reported in detail, intra-rater reliability and internal consistency through Cronbach's alpha. A smaller group of studies were assessed as having moderate quality (Byers et al., 2014; Forman & McKinney, 1978; Gilavand et al., 2016; Kazu & Demirkol, 2014; Reiss & Dyhdalo, 1975; Shamaki, 2015; Solomon & Kendall, 1976). These articles often had rigorous elements to the validity of their design, methods and means of analysis or application of measures of internal consistency and reliability, but not both. Some tended to be based on location-specific assessments of student learning outcomes, while, others did not utilise or report the statistical processes and reliability measures that were evident in studies of strong quality. The remaining articles suffered significant methodological and statistical deficiencies that lowered the quality of their findings. There were correlations between the quality of these studies and the higher incidence of sampling bias.

Table 6: The overall score for the quality of selected studies using the COSMIN 4-point checklist.

FIRST AUTHOR, YEAR	HISTORY	INSTRU-MENTATION	MATURATION	SELECTION	INTERNAL CONSISTENCY	RELIABILITY	OVERALL RESULT
Bottge, B., (2006)	Poor	Good	Poor	Poor	Excellent	Excellent	Low
Chandra, V., (2008)	Good	Good	Good	Fair	Poor	Poor	Low
Cicek, F .G., (2016)	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Strong
Forman, S. G., (1978)	Excellent	Excellent	Excellent	Fair	Fair	Poor	Moderate
Fößl, T., (2016)	Poor	Good	Poor	Poor	Good	Good	Low
McRobbie, C. J., (1993)	Poor	Good	Poor	Poor	Good	Good	Low
Ozerbas, M. A., (2016)	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Strong
Reiss, S., (1975)	Poor	Excellent	Good	Good	Poor	Poor	Moderate
Shamaki, T. A., (2015)	Good	Good	Good	Excellent	Poor	Poor	Moderate
Solomon, D., (1976)	Good	Excellent	Excellent	Fair	Good	Good	Moderate
Tanner, C. T., (2008)	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Strong
Uline, C., (2008)	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Strong
Kazu, I. Y., (2014)	Excellent	Good	Excellent	Good	Fair	Poor	Moderate
Barrett, P., (2015)	Excellent	Excellent	Excellent	Excellent	Good	Good	Moderate
Barrett, P., (2017)	Excellent	Excellent	Excellent	Excellent	Good	Good	Moderate
Byers, T., (2014)	Excellent	Good	Excellent	Good	Excellent	Excellent	Moderate
Chang, C. Y., (2006)	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Strong
Chang, C. Y., (2011)	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Strong
Gilavand, A., (2016)	Excellent	Excellent	Excellent	Excellent	Fair	Excellent	Moderate
Tanner, C. K. (2000)	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Strong

Hiukkavaara School and Community Centre, Lukkaroinen Architects / Ecophon Saint Gobain, Juha Sarkkinen photography.

Discussion

The review sought to establish an evidence base for the connection between learning environment type/s (blended, ILEs, open-plan and traditional) and their impact on student learning outcomes. A descriptive critique is presented based on an analysis of the final selection of 21 studies, centred on the review's three research questions.

What evidence exists that different learning environments have an impact on student learning outcomes?

The systematic review identified a small number of studies that presented empirical evidence of the impact of different learning environments, in particular ILEs, on student academic outcomes. These studies presented evidence that different learning environments, in particular, those aligned with the premise of ILEs, can positively impact student academic achievement. However, studies in the sample that compared open-plan learning environments with traditional classroom spaces suggested that the open-plan setting correlated with a negative impact on student academic achievement. The collective evidence presented by the relatively small number of studies does suggestsome correlation between the design, function and nature of the physical learning environment and learning outcomes.

In this sample, there were only three studies that reported effect sizes. The Fößl et al. (2016) and Bottge et al. (2006) studies reported effect sizes that ranged from d = .31 to d = .78 respectively. However, both had questionable validity and reliability due to the issues of high sampling bias and relatively low quality. Byers et al. (2014), with effect sizes of d = .40 and .41 for English and Mathematics respectively, presented a more valid and reliable assessment of the impact of ILEs in comparison to traditional layouts on student academic achievement. Unlike the Fößl et al. (2016) and Bottge et al. (2006) studies, Byers et al.(2014) utilised a between-group comparison of classes randomly assigned to the different spatial layouts, while controlling for the influence of student cognitive ability. An external, normed measure of non-verbal and verbal reasoning was used as a proxy for student cognitive ability. It showed that students in an ILE outperformed their cognitively matched peers by approximately two grade points (on a 15-point grade scale from A+ to E-) on school-based and moderated tests.

Several studies utilised linear modelling to discern how various designs or physical factors correlated with student academic outcomes but did not report effect sizes. Tanner (2000) identified seven design factors that were found to correlate with student learning outcomes as assessed by the Iowa Test of Basic Skills (ITBS) standardised test. Tanner highlighted that 'compatibility with context', 'clearly defined pathways', 'positive outdoor spaces', 'computers for teachers' and 'positive overall impression' had the greatest positive correlation with increased academic performance on the ITBS from a significant sample (*n* = 22, 678) of students from 44 elementary/primary schools. In a later smaller study (*n* = 1, 916), Tanner (2008) investigated the design pattern effect of movement and circulation (MC), large group meeting places (LGMP), daylighting and views (DLV), and instructional neighbourhoods (IN) on achievement on the ITBS test. In a sample from 24 elementary/primary schools, Hierarchical Linear Modelling (HLM) analysis was able to discern between 2% and 7% additional variance in student achievement on the ITBS, even after controlling for school social and economic status.

In a similar vein to the Tanner studies, Barrett et al. (2015, 2017) sought to establish those specific environmental or physical conditions (i.e. air quality, lighting, noise, temperature, ventilation) that were optimal for student learning in existing classroom spaces. Both Barrett et al. studies focused on the parameters of Naturalness (light, temperature and air quality), Individualisation (ownership, flexibility and connection) and Level of Stimulation (complexity and colour) for a large multi-site sample (*n* = 3, 766). The earlier study (2015) applied correlations of student test results on Overall Progress on the National Curriculum (NC) - Key Stages 1 and 2, against environmental measures (ascertained through hard measures of the environment), and showed all ten parameters were positively correlated with progress. Multilevel (HLM - Two level) modelling portioned between student (highest) and classroom levels. School level variables accounted for little variance (3%). Naturalness parameters (light, temperature and air quality) had a large effect of 28%. Variation between the most effective classroom (Overall progress of 16.05 NC out of 50 points) and the least effective (8.12 NC points), resulted in a 7.93 NC point difference. The impact of the classroom environmental factors, therefore, explains 16% of the variation in students' academic progress. The most recent study (2017), utilised the same data set from the 2015 study but focused on three region-types in the United Kingdom (Blackpool, Hampshire and Ealing). Applying a similar analytical process, the authors ascertained the percentage improvement of 9.3% (Reading), 8.4% (Writing) and 11.7% (Maths) due to physical environmental parameters. Barrett et al. (2017) summarise that for each of the different subject models, the aspects of the classroom environment taken together explained approximately 10% of the variability in student performance.

Another group of studies compared student academic achievement in a blended digital/physical learning environment to that of a traditional classroom setting. Chandra and Lloyd (2008) found that a student sample in the blended learning environment showed improved student achievement (average +5.1 with a decrease in SD of – 3.8) from the pre-testing, whereas, the students in a traditional environment showed a mean post-test decline (average -2.3 with a decrease in SD of – 0.7). Similarly, Cicek and Taspinar (2016) and Kazu and Demirkol (2014) found students in a blended LE outperformed their peers in a more traditional LE with statistically significant ($p < 0.05$) between-group differences. While, Ozerbas and Erdogan (2016) applied two-factor ANOVAs to show a meaningful difference in the academic performance of students in a digital classroom compared to those in a traditional classroom (with no access to digital technology) [$F(1, 56) = 13.041$, $p < .05$]. Collectively, these studies suggest that students who experience a blended LE facilitated by the integration of digital technology achieved relatively higher academic success than their peers in a conventional, traditional classroom environment.

The review sample revealed that changes or improvements to the traditional school learning environment did not always correlate with positive effects on academic achievement. In the comparison of open and traditional classrooms during the 1970's *Open Plan Movement*, Forman and McKinney (1978), Reiss and Dyhdalo (1975) and Solomon and Kendall (1976) found varied trends in student academic performance. All three studies found that students in a traditional classroom outperformed their peers in open-plan spaces on standardised assessment (CAT - California Achievement Test and ITBS test). Forman and McKinney (1978) analysed class means of comparative open and traditional settings for girls and boys. Next, a series of 2 x 2 (Sex x Group) ANOVAs were run using ITBS standard score means and Wallach-Kogan total fluency. Across the comparative sample, students in traditional classrooms showed higher achievement than those in open environments ($F(1, 16) = 7.59$, $p < 0.1$). The composite means of the ITBS for traditional classrooms and open-plan spaces were 57.45 (SD = 6.33) and 47.62 (SD = 11.81) respectively. Through a cluster analysis, Solomon and Kendall (1976) evaluated person-environment interactions showing lower academic achievement in an open-plan space on CAT and ITBS instruments, but with higher assessment in creativity, co-operative behaviour and involvement. Similarly, Reiss and Dyhdalo (1975) found that students in open-spaces were more persistent in their learning, but performed lower on the CAT instrument than those students who remained in the traditional classroom.

What measurement tools have been designed and used for measuring student outcomes in different learning environment types?

Various measurement tools were utilised to evaluate how student learning outcomes differed in the various learning environment types. The studies in the final review with high statistical power achieved through large samples and strong quality and reliability (Barrett et al., 2015, 2017; Chang et al., 2006; Chang et al., 2011; Cicek & Taspinar, 2016; Ozerbas & Erdogan, 2016; Tanner, 2000, 2008; Uline & Tschannen-Moran, 2008) typically utilised assessment through large-scale, standardised assessment instruments. The CAT, ITBS and National Curriculum Key Stages 1 and 2 testing instruments were the most commonly utilised and provided reliable and valid comparative means to determine student progress against the mandated curriculum, often in literacy (i.e. Reading and Writing) and numeracy (i.e. Mathematics). These comparative studies assessed outcomes not typically associated with 21st Century learning. Importantly, the narrative of 21st Century learning underpins the current impetus to reconsider the type and function of school learning environments. However, the most reliable and rigorous studies within this sample do not evaluate how different learning environment types affect the outcomes associated with this 21st Century perspective.

A smaller number of studies, often involving a single research site, utilised existing school-based assessment instruments as a measure of student achievement. Various studies (such as Byers et al., 2014; Chandra & Lloyd, 2008; Cicek & Taspinar, 2016; Kazu & Demirkol, 2014; Uline & Tschannen-Moran, 2008) utilised existing school-based assessment regimes. In these cases, the research context belied the context of the study and the utilisation of existing assessment techniques. Often, the studies did not detail specific features or elements of the nature and type of assessment, which made assessing the internal consistency or reliability through the COSMIN process difficult. However, one study (Byers et al., 2014) outlined the application of a criterion based system, which assessed English and Mathematics subjects based on basic knowledge and procedures through to more applied, complex or open-ended tasks that required higher-order processes.

The interrogation of the sample revealed few measurement tools that were explicitly utilised to measure the incidence and nature of those experiences that best exemplify 21st Century learning. The Fraction of the Cost Challenge (FCC) and Kim's Komet Challenge (KKC) in the Bottge et al. (2006) study focused on assessing student responses to more demanding open-response tests. These tests specifically focused on how students integrated their understanding of two different sets of discrete mathematical concepts to solve a novel problem/ task. The Chang et al. (2011, 2006) studies created and trialled the *Earth Science Classroom Learning Environment* (ESCLE) and *Earth Science Learning Outcomes Inventory* (ESLOI) metrics. Together these tools focus on eliciting students' perceptions on preferred/actual learning environment to their learning achievement and attitude. These studies found, in particular, Chang et al. (2011), that students preferred those environments where both student- and teacher-centred instructional approaches coexisted (Student–Teacher-Balanced Instructional Model - STBIM) over a teacher-centred learning environment (Teacher-Centred Instructional Model). Furthermore, STBIM achievement in and attitude toward earth science were enhanced when the learning environment was congruent with their learning environment preference.

What elements of student learning are quantified by the identified measurement tools?

The dominant use of existing school-based assessment or external standardised testing regimes favoured a particular view of student learning. Often the testing focused on student progression in areas of literacy and numeracy, with a specific focus on standardised assessment. There is little mention, besides Bottge et al. (2006), Byers et al. (2014), Chang et al. (2011) and Cicek and Taspinar (2016), about the influence on student problem-solving skills. However, these studies reported an overall level of achievement, making it difficult to determine if, and to what extent, different learning environment types affected student success in the domain of problem-solving. This lack of evidence is a critical finding of this review. Much of the narrative in the current literature suggests that different spatial layouts (i.e. blended, ILEs and open) are more likely to facilitate 21st Century learning experiences than existing conventional or traditional classroom spaces (Benade; Dovey & Fisher; Dumont & Istance). The sample presented in this review did not present specific measures to assess the impact on the achievement of students as creative and critical thinkers engaged as problem solvers and working in collaboration with peers.

Limitations

The systematic review yielded a small number of studies that met its stated criteria. With only 21 studies included in this sample, such a number would appear to support the assertions made by previous reviews of the literature by Blackmore et al. (2011), Brooks (2011), Gislason (2010) and Painter et al. (2013) that cite a lack of substantive, empirical evidence about the impact of different spatial layouts on student outcomes. Studies that exhibited high statistical power through large samples and strong quality and reliability typically assessed the environmental or physical aspects of the space. These did not discriminate how the specific design, affordance or pedagogical use of ILEs or traditional classrooms impacted student academic outcomes. Their objective lens was unable to deeply explore the nuances of the spaces, their use and contextual factors at the various sites. Furthermore, these studies utilised existing large-scale, standardised assessment instruments. Some argue that such testing regimes promote a type of teaching and learning not aligned with narrative behind 21st Century learning imperatives, which underlie the current interest and investment in different spatial layouts.

The studies that did focus more on a more in-depth comparison of ILE versus traditional classroom spaces often lacked the statistical processes and sampling to genuinely present substantiative empirical evidence. These often singular studies, with fewer incidences of replication in the same context with different student samples, lack the rigour to adequately evaluate the impact of ILEs, or traditional classrooms, on student academic outcomes. Moreover, there were only three studies that reported effect sizes, which is not sufficient for an ensuing meta-analysis. Furthermore, the nature of the methods and means of analysis does support the assertion of Painter et al. (2013) that there are few evaluative methods and metrics currently available to adequately assess the impact of different classroom layouts, ILE or traditional, on student learning outcomes.

Phoenix College , Y2 Architecture, Zac Couyant photography.

Conclusions

The objectives of the review are three-fold. First, to identify what empirical evidence exists that assesses the impact of different learning environment types (blended learning, ILEs, open-plan and traditional) on student learning outcome measures (those with published evidence of reliability and validity). Second, it seeks to identify what measurement tools are used to address this challenge. Finally, it seeks to identify the types of evidence these tools elicit. In other words, what proof exists, how is it being gathered, and what specific knowledge is actually sought? These parameters reflect criticisms regarding our history of learning environment evaluations; namely that they frequently compare 'apples to oranges', have poor sample size and other internal validity issues, utilise tools that bias traditional classroom settings (Gray, 1978; Doob, 1974), and are most often studies conducted three or more decades ago (Imms, in press).

A number of issues are clearly identified in the review.

The paucity of quality evaluation is worrying.

What the review deems robust research on this topic is limited to 21 papers published since 1960, of which only three report a statistical effect. Of these three, only one provides information that comprehensively addresses internal validity issues. This robust systematic review confirms the frequently stated claim that little empirical evidence exists to link student learning outcomes to spatial designs (see, for example, Blackmore, et al., 2011; Brooks, 2011; Cleveland & Fisher, 2014; Gislason, 2010; Painter et al., 2013). In addition, it would appear that in the 2010s we continue to, on occasion, repeat errors from the 1970s where the considerable inconsistency in the design of learning space evaluations means that 'not all studies [could be] considered [methodologically] equal' (Marshall, 1981, p.82).

The studies that met the review's criteria ranged from single-site comparative through to quasi-experimental randomised designs across multiple sites. Their samples sizes ranged widely (17 to 22,000) across primary and secondary schools. The former tended to address student academic outcomes through existing school-based assessment or bespoke measures of creative, critical-thinking, or problem-solving testing instruments. The latter, with significant statistical power and very large samples from multiple schools, assessed the impact of different learning spaces through systemic, standardised literacy and numeracy testing regimes. The review identified a dearth of instruments providing evidence of the impact of ILEs on learning styles deemed characteristic of 21st century needs.

Emerging (but limited) evidence shows a trend that spatial design does positively impact student learning outcomes

Of the studies that meet the criteria for inclusion in this review, inconsistent findings emerge concerning reports of the impact of space on student learning outcomes. Of interest, the range of findings correlates to what might be called a historical ambiguity. Studies from the 1970s that met the review criteria consistently found that students in traditional classrooms out-performed 'open classroom' like-ability peers (for example, Forman & McKinney, 1978; Reiss & Dyhdalo, 1975; Solomon & Kendall, 1976). In contrast, contemporary studies consistently find the opposite; when accounting for student difference, differing pedagogies, and differing classroom typologies, students in (what could be described as) ILEs are found to significantly outperform like-ability peers in a range of key academic subjects (for example see Brooks, 2011; Byers et al., 2014). Space can account for between 7-10% of the variance in academic scores between classes taught in traditional compared to innovative spaces (see, for example, Tanner, 2000 and 2008). The improved 'building-performance' of ILE designs (lighting, acoustics, air quality, etc.) accounts for between 10-16% of the variance in student academic scores compared to traditional designs (for example, see Barrett et al. 2015 and 2017). Blended learning environments (technology plus ILEs) account for a statistically significant improvement in student academic scores (Chandra & Lloyd, 2008; Cicek & Taspinar, 2016; Kazu & Demirkol, 2014; Ozerbas & Erdogan, 2016). It must be stressed that this growing evidence is limited in scope; as an indicator, there are insufficient studies that report an effect size, to warrant any meta-analysis on this topic.

An issue is that historically, the evaluation tools utilised do not always measure the learning characteristics ILEs were designed to achieve

Evaluations of learning spaces from the 1970s consistently failed to define key concepts; they were predominately based on 'snapshots' rather than longitudinal designs, and mostly used low sample sizes (Gray, 1978). Their irregularity of findings indicated poor consistency between measures of 'openness' and outcome variables (Jackson, 1980) with, in some cases, the impact of 'open learning' being assessed despite the fact it occurred in traditional classroom settings (McPartland & Epstein, 1977). During this era, the tools being used to quantitatively measure the impact of open learning designs were often designed for traditional settings and were unable to capture data on the characteristics of student learning that were the predominant driver of good open space design (Imms, Cleveland & Fisher, 2016).

This review suggests a similar phenomenon occurring in contemporary evaluations, but now mediated by a subsequent three decades of research that has provided more nuanced understandings of the significance of surface-to-deep learning. A cautionary finding from this review is that recent evaluations of the impact of learning space design on student learning outcomes rarely focus on students' learning processes. They consistently utilise either large-scale standardised testing or smaller school-based assessments that favour particular types of student learning – most often measures of literacy and numeracy. The benefit here is the validity very large sample sizes provide. The caution is regarding attributing meaning to results from these approaches. On balance, large statistically powerful studies of this type lack space-specific items particular to 21st century learning – the very issue ILE designs were asked to accommodate. In comparison, the review's selection process (in particular its use of the Cochrane Collaboration Tool) illustrates how smaller ILE-bespoke studies that can address 21st century learning characteristics are prone to criticisms of internal validity and reliability.

An evidence-based 'narrative' is emerging but requires considerable high-quality research to substantiate

This does not mean a narrative is not emerging; there is a growing body of research that may eventually prove ILE's worth beyond just literacy/ numeracy measures. A finding of this review is that, of the few studies considered 'robust', some have evolved methodologies that accommodate affective domain learning characteristics sought by the ILE concept. Indeed, even in the 1970s a wealth of qualitative research argued that open designs provide students marked improvement in the student-centric learning approaches that develop collaborative, innovative, creative learning skills (Doob, 1974). In more recent years some innovative methodological experimentation has added 21st century learning characteristics as measurable items in sound quantitative studies. Bottge et al. (2006), Chang et al. (2006) and Chang et al. (2011) addressed the deep learning issue and included data that showed student preference for learning environments that facilitated student-centered learning approaches over the teacher-centric classrooms. Byers et al. (2014) have developed the Single Subject Research Design (SSRD) approach, borrowed from the applied health sciences, to isolate space as a variable and report confidence intervals on perceptions of engagement and teaching performance.

This review finds that very few quality evaluations exist regarding the impact of ILEs on student learning outcomes. It finds an historical disjuncture, with two 'eras' of evaluation (1970s versus the 2000s) differing in terms of primary findings. It finds that methodological weaknesses accounts for this difference, with recent studies enjoying access to larger sample sizes and advanced evaluation methods. It finds the focus of ILE evaluations is now accommodating to a greater degree the qualities of space that have driven ILE development (the acquisition of 21st century learning skills), but still remains methodologically problematic. It finds a trend is becoming evident that suggests ILEs have a positive impact on student learning outcomes. While this is optimistic, it cautions over-stating of this trend at this time.

References

Alterator, S., & Deed, C. (2013). *Teacher adaptation to open learning spaces. Issues in Educational Research, 23*(3), 315-330. Retrieved from http://www.iier.org.au/iier23/ alterator.html.

Barrett, P., Davies, F., Zhang, Y., & Barrett, L. (2015). The impact of classroom design on pupils' learning: Final results of a holistic, multi-level analysis. *Building and Environment*, *89*, 118-133. doi:10.1016/j.buildenv.2015.02.013.

Barrett, P., Davies, F., Zhang, Y., & Barrett, L. (2017). The holistic impact of classroom spaces on learning in specific subjects. *Environment and Behavior, 49*(4), 425-451. doi:10.1177/0013916516648735.

Benade, L. (2017). Is the classroom obsolete in the twenty-first century? *Educational Philosophy and Theory, 49*(8), 796-807. doi: 10.1080/00131857.2016.1269631.

Blackmore, J., Bateman, D., O'Mara, J., & Loughlin, J. (2011). *Research into the connection between built learning spaces and student outcomes: Literature review*. Melbourne: Victorian Department of Education and Early Childhood Development. Retrieved from http://www.education.vic.gov.au/Documents/about/programs/infrastructure/ blackmorelearningspaces.pdf.

Bottge, B., Rueda, E., & Skivington, M. (2006). Situating math instruction in rich problem-solving contexts: Effects on adolescents with challenging behaviors. *Behavioral Disorders, 31*(4), 394-407. doi:10.1177/019874290603100401.

Brooks, D. C. (2011). Space matters: The impact of formal learning environments on student learning. *British Journal of Educational Technology, 42*(5), 719-726. doi:10.1111/j.1467-8535.2010.01098.x.

Byers, T., Imms, W., & Hartnell-Young, E. (2014). Making the case for space: The effect of learning spaces on teaching and learning. *Curriculum and Teaching, 29*(1), 5-19. doi:10.7459/ct/29.1.02.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research on teaching.* Chicago: Rand McNally.

Chandra, V., & Lloyd, M. (2008). The methodological nettle: ICT and student achievement. *British Journal of Educational Technology, 39*(6), 1087-1098. doi:10.1111/j.1467-8535.2007.00790.x.

Chang, C. Y., Hsiao, C. H., & Barufaldi, J. P. (2006). Preferred - Actual learning environment "spaces" and earth science outcomes in Taiwan. *Science Education, 90*(3), 420-433. doi:10.1002/sce.20125.

Chang, C. Y., Hsiao, C. H., & Chang, Y. H. (2011). Science Learning Outcomes in Alignment with Learning Environment Preferences. *Journal of Science Education and Technology, 20*(2), 136-145. doi:10.1007/s10956-010-9240-9.

Cicek, F. G., & Taspinar, M. (2016). *Laboratory Control System's Effects on Student Achievement and Attitudes. Eurasian Journal of Educational Research, 64*, 247-264. doi: 1610.14689/ejer.2016.64.14.

Cleveland, B. & Fisher, K. (2014). The evaluation of physical learning environments: A critical review of the literature. *Learning Environments Research,17*, 1-28.

Doob, H. (1974). *Summary of research on open education: An ERS Research Brief.* Arlington, VA: Educational Research Service.

Dovey, K., & Fisher, K. (2014). Designing for adaptation: The school as socio-spatial assemblage. *The Journal of Architecture, 19*(1), 43-63. doi:10.1080/13602365.2014.882376.

Dumont, H., & Istance, D. (2010). Analysing and designing learning environments for the 21st century. In Dumont, H., Istance, D. & Benavides, F. (Eds.), *The nature of learning: Using research to inspire practice*, 19-34. Paris: OECD publications. doi: 10.1787/9789264086487-en.

Forman, S. G., & McKinney, J. D. (1978). Creativity and achievement of second graders in open and traditional classrooms. *Journal of Educational Psychology, 70*(1), 101-107. doi:10.1037/0022-0663.70.1.101.

Fößl, T., Ebner, M., Schön, S., & Holzinger, A. (2016). A field study of a video supported seamless-learning-setting with elementary learners. *Journal of Educational Technology & Society, 19*(1), 321-336.

Gilavand, A., Espidkar, F., & Gilavand, M. (2016). Investigating the impact of schools' open space on learning and educational achievement of elementary students. *International Journal of Pediatrics, 4*(4), 1663-1670. doi:10.22038/ijp.2016.6672.

Gislason, N. (2010). Architectural design and the learning environment: A framework for school design research. *Learning Environments Research, 13*(2), 127-145. doi:10.1007/s10984-010-9071-x.

Gray, W. (1978). Open areas and open education re-examined: A research study. *Australian Journal of Teacher Education, 3*(1).

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., … Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928. doi:10.1136/bmj.d5928.

Imms, W. (in press). Innovative learning spaces: Catalysts/agents for change, or 'just another fad'? In S. Alterator & C. Deed (Eds.). *School space and its occupation: The conceptualisation and evaluation of new generation learning spaces.* Sense Publishers, Netherlands.

Imms, W., Cleveland, B., & Fisher, K. (2016). Pursuing that elusive evidence about what works in learning environment design. In Imms, W., Cleveland, B., & Fisher, K. *Evaluating learning environments* (pp. 3-17). Rotterdam: Sense Publishers.

Jackson, G. (1980). Methods for integrative reviews. *Review of Educational Research, 50*, 438-460.

Kazu, I. Y., & Demirkol, M. (2014). Effect of blended learning environment model on high school students' academic achievement. TOJET: *The Turkish Online Journal of Educational Technology, 13*(1), 78-87.

Marshall, H. (1981). Open classrooms: Has the term outlived its usefulness? *Review of Educational Research, 51*(2), 181-192.

MCEETYA. (2008). Melbourne declaration on educational goals for young Australians. Victoria, Australia: Ministerial Council on Education, Employment, Training and Youth Affairs.

McPartland, J. & Epstein, J. (1977). Open schools and achievement: Extended tests of a finding of no relationship. *Sociology of Education, 50(2),* 133-144.

McRobbie, C. J., & Fraser, B. J. (1993). Associations between student outcomes and psychosocial science environment. *The Journal of Educational Research, 87*(2), 78-85. doi: 10.1080/00220671.1993.9941170.

New Zealand Ministry of Education. (2011). The New Zealand school property strategy 2011-2021. Wellington, New Zealand: New Zealand Ministry of Education. ISBN: 978-0-478-36794-2 (Web). Retrieved from: http://gdsindexnz.org/wp-content/ uploads/2015/10/The-New-Zealand-School-Property-Strategy-2011-2021.pdf.

New Zealand Ministry of Education. (2014). The Ministry of Education's statement of intent 2014-2018. Wellington, New Zealand: New Zealand Ministry of Education. Retrieved from https://www.education.govt.nz/assets/Documents/Ministry/Publications/ Statements-of-intent/2014SOI.pdf.

OECD. (2013). Innovative Learning Environments. Paris: Organisation for Economic Cooperation and Development. doi: 10.1787/9789264203488-en.

OECD. (2015). Schooling redesigned: Towards innovative learning systems. Paris: Organisation for Economic Cooperation and Development. doi:10.1787/9789264245914-en.

Ozerbas, M. A., & Erdogan, B. H. (2016). The Effect of the Digital Classroom on Academic Success and Online Technologies Self-Efficacy. *Journal of Educational Technology & Society, 19*(4), 203-212.

Painter, S., Fournier, J., Grape, C., Grummon, P., Morelli, J., Whitmer, S., & Cevetello, J. (2013). *Research on learning space design: Present state, future directions.* Society of College and University Planning.

Reiss, S., & Dyhdalo, N. (1975). Persistence, achievement, and open-space environments. *Journal of Educational Psychology, 67*(4), 506-514.

Shamaki, T. A. (2015). Influence of Learning Environment on Students' Academic Achievement in Mathematics: A Case Study of Some Selected Secondary Schools in Yobe State-Nigeria. *Journal of Education and Practice, 6*(34), 40-44.

Solomon, D., & Kendall, A. J. (1976). Individual characteristics and children's performance in open and traditional classroom settings. *Journal of Educational Psychology, 68*(5), 613-625. http://dx.doi.org/10.1037/0022-0663.68.5.613.

Tanner, C. K. (2000). The influence of school architecture on academic achievement. *Journal of Educational Administration, 38*(4), 309-330. doi:10.1108/09578230010373598.

Tanner, C. K. (2008). Explaining relationships among student outcomes and the school's physical environment. *Journal of Advanced Academics, 19*(3), 444-471. doi: 10.4219/jaa-2008-812.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research, 21*(4), 651-657. doi: 10.1007/s11136-011-9960-1.

Uline, C., & Tschannen-Moran, M. (2008). The walls speak: the interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration, 46*(1), 55-73. doi:10.1108/09578230910955818.

Wall, C. (2009). *Building the Education Revolution: National coordinator's implementation report (February-September 2009)*. Canberra: Australian Federal Government. Retrieved from http://apo.org.au/node/19434.

Our partners

THE UNIVERSITY OF MELBOURNE

MONASH University

NSW GOVERNMENT | Education

ACT Government Education

Catholic Education Diocese of Parramatta

Churchie.

Queensland Government

MARSHALL DAY Acoustics

ASSOCIATION FOR LEARNING ENVIRONMENTS
Enhancing the Educational Experience
Australasia

hayball

Steelcase® EDUCATION

AUSTRALIAN SCIENCE & MATHEMATICS SCHOOL

WOODLEIGH school
A SCHOOL of INDEPENDENT THOUGHT

MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga
New Zealand Government
Te Kāwanatanga o Aotearoa

Ecophon® SAINT-GOBAIN
A SOUND EFFECT ON PEOPLE

MAAS Museum of Applied Arts & Sciences