DCU
Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

REPORT 2 ANNEX

Weaving the Literature on Integration, Pedagogy and Assessment: *Insights for Curriculum and Classrooms*

Patrick Burke
Paula Lehane

This annex accompanies the following report:

Weaving the Literature on Integration, Pedagogy and Assessment: *Insights for Curriculum and Classrooms*

Report 2
Examining Integration, Pedagogy and Assessment in the Context of the Redeveloped Irish Primary School Curriculum

Dr Patrick Burke
Dr Paula Lehane

Institute of Education
Dublin City University

May 2023

This annex can be cited as:
Burke, P. & Lehane, P. (2023). *Weaving the Literature on Integration, Pedagogy and Assessment: Insights for Curriculum and Classrooms – Annex 2*. Dublin: National Council for Curriculum and Assessment.

Table of Contents

Section 1

Curriculum Integration Studies

Section 1

Curriculum Integration Studies

The table on the following accompanies Chapter 2 of Report 2.

Lead Author (Year) Location	Intervention	Study Design Sample	Outcomes	Study Limitations
Akbar (2012) United States	The teacher in this study integrated English Language Arts (ELA) concepts into mathematics; vocabulary instruction, use of comprehension strategies, writing about mathematical concepts. Student understanding was tested before and after using the chapter test from the classroom textbook; further data was also gathered through questionnaire/Likert items and analysis of student work and teacher lesson plans.	Single Group *N*=64 (5th Grade)	Scores in maths increased from pre to post-test. Students reported positive attitudes towards this form of mathematics teaching/ integration.	(i) Intact groups (ii) There is a lack of clarity in the reporting of the intervention/control groups' scores, meaning it is difficult to be confident in the interpretation of these data.
Alghamdi (2017) Saudi Arabia	This study examined the effects of an integrated mathematics and science curriculum on academic achievement in a private Saudi Arabian elementary school. An integrated unit was developed from a 5th grade 'Sound and Light' science unit and a 5th grade 'Perimeter, Area and Size' mathematics unit. The control group addressed each subject's unit separately while the treatment group experienced an integrated approach to instruction.	Between Groups *N*=162 (5th Grade) Treatment: *n*=76 Control: *n*=86	There was a significant difference ($p < 0.01$) between the post-test science and maths scores of the treatment and comparison groups. While all groups had improved test scores, the treatment group outperformed the comparison group on both tests (Science: 0.44 ES, Maths: 0.49 ES).	(i) Suitability of the outcome measures involved should be queried in terms of its suitability i.e. was a vocabulary test the best way to measure learning? (ii) Further details on the integrated science/maths unit was required to fully understand the differences between the experimental conditions (iii) Inadequate reporting of statistical analyses and measures (iv) No mention of observation data gathered
An (2014) United States	An experimental approach (using a pretest-posttest control group design) was used in this study to examine the impact of music-mathematics integrated lessons on 3rd grade students' mathematics achievement and dispositions, including beliefs about success, attitude, confidence, motivation, and	Between Groups *N*=56 (3rd Grade) Treatment: *n*=28 Control: *n*=28	Analysis of the post-test scores indicated that the treatment group ('music') had statistically significantly higher scores in mathematical dispositions than the control group ('non-music')	(i) Research did not seem to address the unequal role assigned to the two subjects i.e. maths learning was prioritised over music learning (ii) Music based outcome measures were required to support some of the assertions made in the discussion section (iii) Possible novelty effect

	usefulness. The students in the treatment group received music-mathematics integrated lessons, while the students in the control group received a more traditional approach to mathematics instruction.		students after the intervention (p<0.001; d=1.37). Post-hoc tests revealed that the treatment group demonstrated a statistically significant higher disposition score in the post-test than pre-test in all six mathematics disposition areas measured with large to very large effect sizes noted; mathematics confidence [d=0.94], attitude [d=1.46], usefulness [d=1.88], success [d=1.18], motivation [d=0.74], and beliefs [d=1.98].	(iv) Minimal discussion of how potential confounders were controlled for
An (2013) United States	This study investigated how two elementary school teachers designed music-mathematics interdisciplinary lessons (1st Grade, 3rd Grade). The two participant teachers attempted to design and implement music activities as an integrated part of their regular mathematics lessons over a five-week period.	Single Group N=46 (n=21; 1st Grade, n=25; 3rd Grade)	The ability levels of the first and third grade students in all three mathematical areas as assessed by the researcher designed rubric (which aligned with that state's core standards) showed statistically-significant improvements. Large effect sizes were found in both 1st and 3rd grade students before and after the intervention across the 'model-strategy-application' elements of the rubrics: Model: d=1.66, d=3.40 Strategy: d=1.75, d=3.44 Application: d=1.70, d=3.00	(i) Bias within sample (teacher qualifications/ interest) (ii) Researcher designed instruments (iii) Short intervention period (Hawthorne effect) (iv) Minimal discussion of how potential confounders were controlled for

Atalay (2015) Turkey	Study examined the effects of an 'Integrated Curriculum Model' in a social studies classroom for gifted and talented learners. One class experienced an integrated approach to instruction involving a range of pedagogical methods, the other experienced a 'standard' approach instruction involving text-books and lecture style teaching methods.	Between Groups *N*=21 (4th Grade; Gifted learners) Treatment: *n*=11 Control: *n*=10	Social Studies Achievement Test: Achievement was significantly better in the treatment condition. Cornell Critical Thinking Test: Overall achievement was significantly better in the treatment condition. Torrance Test of Creative Thinking: Overall performance in this test was better in the treatment condition.	(i) Researcher delivered instruction to the experimental group (ii) Sample size and composition (iii) Control group's programme is poorly defined
Bergen-Cico (2015) United States	This study examined the impact of integrating yoga into ELA classes. Specifically, it compared scores on a measure of student self-regulation across an intervention group (yoga embedded ELA) and control group (which included some instruction on mindfulness, but not regular yoga).	Between Groups *N*=144 (6th Grade) Treatment: *n*=72 Control: *n*=72	Students in the treatment group (yoga embedded ELA) had significantly higher self-regulation scores at mid-year and year end, despite having similar baseline scores (partial eta-squared ES=.03). Qualitative teacher feedback indicated positive effects. 60% of students reported that they found yoga helpful; stating it helped to calm them down and improve concentration. The remaining 40% of negative responses pointed out that it led to a loss of instructional time; that it was not as effective at	(i) Intact groups (ii) Statistical analysis does not fully account for the use of intact groups (iii) Individual differences in teachers (between the control/intervention group) may account for differences in scores - this is not accounted for (iv) Reliance on self-report data (v) Potential Hawthorne effects (v) Relatively small sample size

			helping relaxation as envisaged or indeed that it caused them to feel *less* regulated. Parents were broadly positive about yoga integration.	
Birsa (2018) Slovenia	This study presented the key findings arising from a piece of experimental research that examined the effects of teaching visual arts concepts (specifically sculpture) using cross-curricular integration approaches.	Between Groups *N*=274 (5th Grade) Treatment: *n*=160 Control: *n*=144	Students were observed and assessed under the following headings of a researcher designed rubric: motivation, creativity in solving the four sculpting tasks, and art concept comprehension (written examination/art products). According to the authors, a discriminant analysis revealed that the experimental group made more progress than the control group in creative engagement in the implementation of the sculpting tasks, as reflected in the grades achieved by students in sculptural works - this was the greatest difference in scores noted between the treatment and control groups. Measures related to student motivation indicated that students in the experimental group were more motivated than their peers in the control group.	(i) Inadequate discussion on how a range of potential confounders/biases were controlled for (ii) Suitability of outcomes measures (iii) Inadequate reporting of statistical analyses (e.g. assumptions)

Bravo (2014) United States	This study adopted a quasi-experimental approach to determine how science-literacy integration benefits English learners. Children in the intervention group took part in a series of 40 sessions built around a unit on space, balancing first-hand/inquiry approaches and literacy-focused activities. Children in the comparison group took part in science lessons based on the same content. Pre-post test data was gathered.	Between Groups *N*=115 (4th/5th Grade) Treatment: *n*=77 Control: *n*=95	On the basis of observational data, the researchers conclude that linguistic development was better supported in the environment of intervention group classes, due to an increased focus on oral communication and sense making. Students in the intervention group had statistically significant higher scores in science understanding and vocabulary. There was no difference in science reading comprehension. This pattern was replicated when focusing only on the outcomes for English language learners.	(i) The authors identify homogeneity in the languages of the EL (mainly Spanish) as a limitation and the limited amount of time spent observing. (ii) Little attention is given to the curricular implications of integrating literacy in this way (iii) Query if more advanced statistical modelling may have better accounted for the findings presented, rather than ANOVA analysis
Brugar (2012) United States	This study took place in a curricular context (US) in which social studies and visual arts were less likely to be taught due to testing requirements in literacy and numeracy. Two teachers implemented an interdisciplinary history-literacy-visual arts unit; student outcomes were compared to those in a comparison class. Pre-post data was gathered.	Between Groups *N*=50 (5th Grade) Treatment: *n*=27 Control: *n*=23	All students experienced improvements in learning regardless of condition. Students in the experimental group made greater gains, statistically speaking (*d*=0.84), than the students in the comparison group (*d*=0.52) between the pre- and the post-assessment Analysis of the researcher-designed assessment demonstrated that students scored significantly higher in the	(i) Intact groups (ii) High level of support provided in instructional materials (iii) Potential research bias in evaluation of their own materials (iv) Use of researcher-designed and unvalidated measures (v) Sample size

			treatment classes than in the comparison class after the intervention; *d*=1.02. Qualitative observational data indicated that students demonstrated both substantive and procedural engagement during interdisciplinary lessons. Students demonstrated aesthetic development (relating to visual arts integration) and historical thinking at a variety of levels in their classroom contributions. Teachers noted a number of benefits: (i) student engagement; (ii) promotion of higher order thinking skills and (iii) teacher learning (e.g. their own interest and ownership grew over time). boundaries in timetabling)	
Bryant (2012) United States	Students in the treatment group of this study had their fluency instruction supplemented with musical strategies over the course of eight weeks, e.g. singing, clapping, body percussion, found instruments. Other music-integrated fluency instruction included the use of ascending/descending scales to match speech intonation patterns (e.g. to signal questions). Students in the control group also received fluency instruction, but without the	Between Groups *N*=115 (1st Grade) Treatment: *n*=55 Control: *n*=60	Comparison of post-test scores indicated that there was no significant difference between treatment and control groups on the DIBELS nonword fluency measure (*d*=0.27) but that there was a significant difference for the DIBELS phoneme segmentation fluency measure (*d*=0.747 (favouring treatment students).	(i) Intact groups (ii) No mention of the music curriculum - very large focus on the literacy curriculum (iii) No measure of musical outcomes adopted (iv) No observation/reliance on teacher self-report of fidelity measures (v) The statistical analyses do not account for nested data. (vi) Limited conceptual explanation for the significant/non-significant findings [beyond pointing to issues with statistical power]

	musical elements. Pre-post data was gathered.			
Byrd (2019) United States	This study compared the achievement of fifth grade students in two conditions: arts integrated and traditional instruction. The study draws on qualitative and quantitative measures, each of which provide conflicting findings. Pre-post data was gathered.	Between Groups *N*=180 (1st Grade) Treatment: *n*=94 Control: *n*=95	From a qualitative perspective, a number of benefits of arts integration were highlighted; teachers stated that it supported student engagement/excitement and reduced behavioural issues; this was supported by observational data. No statistically significant difference was found between arts-integrated and traditional instruction groups in the following measures: GPA gain scores, 6+1 Traits of Writing scores, ELA grade gain scores.	(i) Statistical analyses pay limited attention to the use of multiple comparisons and nested data. (ii) The author suggest that a longer time period may have given rise to statistically significant differences. (iii) Limited information on the methods/approaches being used in both conditions. (iv) Lack of clarity around role of researcher in supporting teachers.
Cannon-Ruffo (2020) United States	This study compared the cognitive and affective outcomes of learners' participation in an integrated STEM curriculum involving educational robotics. The intervention consisted of 10-12 sessions and used Lego MindStorms EV3 robotics kits. Progress was measured using a researcher-designed STEM knowledge assessments and the affective scales from the Partnerships in Education and Resilience (PEAR) Common Instrument Suite.	Between Groups *N*=80 (4th Grade) Treatment: *n*=45 Control: *n*=35	Results indicated that the intervention was associated with higher STEM achievement (Cohen's *d*=1.12) and perseverance (Cohen's *d*=0.45). There were no statistically significant gender main effects or interaction effects for either cognitive or affective outcomes.	(i) Intact, non-equivalent groups (ii) Short duration of intervention (<3 weeks) (iii) Some statistical analyses may not have been appropriate to conduct given the violations of certain assumptions
Casady (2015) United States	This study examined differences in the achievement of fourth graders taught in an experimental interdisciplinary condition and a more traditional, subject-based condition.	Between Groups *N*=53	T-test analysis demonstrated that science scores post-intervention were significantly higher in the integrated condition (p<.001).	(i) Intact groups (ii) Science scores were not gathered prior to the intervention - it cannot be ascertained if higher

	Teaching took place over the course of one school year (August to May). The researcher was the classroom teacher for the interdisciplinary condition. Pre-post data was gathered using as a district-constructed assessment of Social Studies and a standardised assessment of English reading comprehension and vocabulary (STAR Renaissance instruments). Student questionnaires and administrator interviews conducted before and after the intervention. Post-intervention data gathered in relation to science (district constructed assessment).	Treatment: *n*=38 Control: *n*=15	ANOVA analysis of social studies scores indicated significant differences (p<.001), with the integrated condition showing greater growth. ANOVA analysis of ELA scores indicated that there was no significant difference between groups on any measures. Analysis of questionnaire responses pre and post led the author to conclude that students in the integrated condition had a greater appreciation for social studies; administrators were generally positive about their experiences.	scores in the treatment group can be attributed to the intervention (iii) Researcher effect cannot be ruled out - the gains in the integrated condition may be attributable to teacher differences rather than mode of delivery differences. (iv) Qualitative analysis is quite superficial. (v) Potential for researcher bias in reporting of some of the qualitative data (e.g. relationship with administration may affect the answers provided).
Cecchini (2020) Spain	This quasi-experimental study examined the effects of an interdisciplinary educational approach integrating physical education and mathematics on light and moderate–vigorous physical activity (PA), sedentary behaviour, and learning subtraction. Pre-post data was gathered using a subtraction test developed by Yánez & Bethencourt (2004) and PA data from GT3X Activity monitors.	Between Groups *N*=46 (1st Grade) Treatment: *n*=23 Control: *n*=23	Students from the treatment group reached higher levels of light PA (d=2.97), moderate–vigorous PA (d=2.35), and spent less time in sedentary behaviour (d=4.01), than students who attended regular classroom lessons. Moreover, the students from the treatment group achieved higher scores in subtraction learning than the control group (d=1.20).	(i) Intact groups (ii) Relatively short intervention period (3 weeks), (iii) Generalisability concerns (e.g. subject specialist rather than generalist teachers were involved, sample size)

Cervetti (2012) United States	In this experimental study, students/teacher undertook a model of integrated science and literacy instantiated in the topic of light. Students/teachers in the control group implemented 'business as usual' teaching of similar content. Students engaged in work that balanced both first-hand inquiry and literacy skills about the topic. Outcomes were measured using researcher-designed measures of science understanding; science writing; vocabulary; reading comprehension. All of the measures involved underwent strict piloting/validation procedures,	Between Groups *N*=94 (4th Grade Classrooms) Treatment: *n*=47 classrooms Control: *n*=47 classrooms	Moderate effect sizes for science learning/understanding (ES=.65); small effect size for vocabulary (ES=.22); moderate multivariate effect for writing dimension (ES=.40). No difference in performance was noted between groups on measures of reading comprehension.	(i) Short-term-effects only measured over course of one unit. (ii) Self-report of fidelity measures. (iii) Highly scaffolded integration that may not be replicable by teachers on their own
Chand O'Neal (2017) United States	This large-scale study examined the effects of arts integration on student creativity, engagement and academic outcomes after one year. The study involved a comparison of learners involved in the Changing Education through the Arts (CETA) program with children at matched control schools using traditional classroom instruction (quasi-experimental design with multigroup analyses). Affective outcomes (Runco Creativity Assessment Battery; Chand-O'Neal & Schulz Begle Student Engagement Survey) were examined at the student level (self-report), parent level (parent perceptions of their child's creativity and engagement in schoolwork), and teacher level (teacher perceptions of their student's creativity and engagement in schoolwork). Standardised	Ex-Post Facto Design *N*=746 (4th and 5th Grade) *N*=746 (Parents) *N*=86 (In-Service Teachers) Treatment: *n*=552 learners Control: *n*=194 learners	Key findings demonstrated that overall *Attitudes about Art* did not show a significant change for either group but students receiving arts integrated instruction reported a greater increase in one subscale (positive attitudes about artists; *d*=0.16) over the course of the study compared to matched controls. Students receiving arts integrated instruction reported little change in Engagement subscales (Interest, Effort, and Challenge) over the course of the study while students who did not receive arts integrated instruction showed an	(i) Unequal group sizes (ii) Pre-test scores indicated that baselines between schools were very different on a range of measures (iii) State-Based standardised tests can be problematic to interpret (iv) Students had been involved in CETA since Kindergarten and the cumulative effects of that were not clearly elicited in the research.

	Test Scores in Reading and Math were also examined.		increase in their Interest (*d*=0.25) and Effort (*d*=0.35) scores over time, and a decrease in Positive Challenge (*d*=-0.37) scores over time. No significant differences were found in Math and Reading standardised test scores between the treatment and control groups. The author noted that, overall, the changes in standardised test scores were comparable for both groups. Teachers in arts integrated schools reported a greater increase in their students' effort and engagement (*d*=0.29) as well as their overall performance in the Teacher Evaluations of Student Creativity (*d*=0.24). There were no significant differences for parent reports of student engagement or creativity between treatment and control groups across time.	
Cotič (2021) Slovenia	This non-randomised, quasi-experimental study examined the effects of an integrated, cross-curricular approach to teaching mathematics and science with experiential	Between Groups *N*=304 (3rd Grade)	No statistically significant differences between the treatment and control groups on items	(i) Intact groups (ii) short intervention (3 hours) (iii) Researcher designed materials

	learning techniques on the knowledge of 3rd grade primary school pupils in Slovenia. Pre-Post data gathered using researcher-designed test materials that were aligned with TIMSS domains.	Treatment: *n*=149 Control: *n*=155	related to factual knowledge were noted. Regard the domains of 'conceptual understanding' and 'reasoning and analysis', statistically significant differences were identified for both. On both domains, students in the treatment condition outperformed the control condition.	(iv) Intervention was poorly described - difficult to determine what features it included beyond the term 'experiential learning'
Cunnington (2014) United States	This study investigated the impact of the 'Framing Student Success: Connecting Rigorous Visual Arts, Math and Literacy Learning' project, which delivered interdisciplinary visual arts, maths and literacy instruction in New York City Title 1 (high poverty) schools. A cluster randomised trial was conducted across six schools (three treatment, three control), in which artists/instructors taught integrated lessons in classrooms with support from teachers. The study reports on academic and other outcomes. Outcomes examined included: document analysis, observation in classrooms with a locally developed rubric, observation of Professional Development sessions, focus groups/interviews/surveys with relevant parties, New York State ELA and math standardised test scores, rubric to assess artwork, Visual Arts Benchmark Arts State Assessment, Studio Habits of Mind	Between Groups N=6 schools Treatment: *N*=3 schools *n*=545 students *n*=66 teachers *n*=3 visual arts specialists *n*=15 administrators Control: *N*=3 schools *n*=456 students *n*=52 teachers *n*=3 visual arts specialists *n*=15 administrators	Multiple regression analyses held constant the standardised test scores from the preceding year, dosage, and demographic characteristics; they revealed that treatment group students made greater learning gains than those in the control group in both ELA (ES=0.12) and maths (ES=0.24). Repeated-measure ANOVA revealed that students in the treatment group showed significantly greater growth in reflecting (ES=0.33). There was no significant difference in Studio habits of Mind; Chi-square analyses of Benchmark Arts Assessments showed that control and treatment	(i) Variation in dosage amongst treatment students/schools (ii) limited information provided on fidelity of implementation;

	rubric (assesses critical thinking/approaches to study/learning).		groups performed similarly but significantly larger portions of the treatment reached proficiency on seven VA skills.	
Doyle (2014) United States	This paper briefly summarises the results of the first two years of a three-year professional development programme that supports teachers with arts-integrated approaches to teaching and learning (Collaboration: Teachers and Artists; CoTA). A range of data was gathered to evaluate the programme including teacher interviews and student standardised assessment data (state ELA assessments). A quasi-experimental multi-site design research study with pre-post data collection was conducted.	Between Groups N=6 Elementary Schools Treatment: N=3 schools Control: N=3 Schools	Student benchmark testing results in the first year were mixed. At the sixth-grade level, CoTA students showed significantly lower gains than did comparison (partial η2=.035). While at the second and third grade levels, CoTA students showed significantly greater gains on the tests (partial η2=.061, partial η2=.017). For Year 2, an ANCOVA found that there was no statistically significant difference between CoTA and comparison student growth in grades 3, 4, or 6. However, subsequent analyses that employed covariate controls, found a statistically significant difference between CoTA and comparison students' scores in second and fifth grade with higher adjusted means for CoTA students.	(i) Intact groups (ii) Study described outcomes at the end of Year 2 (rather than Year 3) (iii) The analysis applied for both qualitative and quantitative data needed a more detailed description i.e. it appears that the comparison and treatment schools may have been significantly different in terms of student functioning (CoTA explained between 2% and 6% of the variance, while intervention-comparison differences on the pre-test accounted for about 50% of the variance) (iv) Generalisability (i.e. intensive PD programme)
Duke (2021) United States	In this cluster randomized controlled trial, pairs of teachers in each school were randomly assigned to either the treatment	Between Groups	Using hierarchical linear modelling to control for gender, race/ethnicity, mother's level of	(i) Use of researcher-developed assessment instruments.

	group or the comparison group (which involved teaching their regular, non project-based social studies curriculum); baseline equivalence was confirmed. Teachers in the experimental group taught four integrated social studies/literacy units addressing economics, geography, history and civics/government. Outcome measures included: social studies assessment [researcher designed and validated]; informational reading assessment [researcher designed and validated]; informational writing assessment [researcher designed and validated]; motivation survey for social studies, literacy learning, integrated social studies and literacy [based on other available surveys, but researcher-designed]; structured observations; interviews. No pre-intervention data was gathered.	*N*=48 (2nd Grade classrooms) *N*=684 (2nd Grade students) Treatment: *n*=289 Control: *n*=395	education and baseline scores, the experimental group scored higher on social studies knowledge (ES=0.48) and informational reading (ES=0.18). There was no significant difference for informational writing or motivation measures. Teachers that enacted the units with higher measures of fidelity with improved scores on all measures: social studies (ES=0.27), reading (ES=0.58), writing (ES = 0.24), and motivation (ES =0.29).	(ii) Overall sampling of schools was not random (convenience) (iii) High level of instructional support not replicated easily; (iv) Limited qualitative/ experiential data reported in this study
Fazio (2019) Canada	This study employed a design-based research approach to track how integrated science and literacy teaching would impact on student scores on vocabulary, comprehension and science content. Five teachers and the children in their classrooms participated. To measure student outcomes, standardised measures of vocabulary and reading comprehension (from the norm-referenced Canadian Achievement Test) were administered alongside a researcher designed assessment of science content/knowledge. Pre-post test data was gathered. A range of	Single Group N=5 classrooms N=118 (5th Grade)	Student growth in vocabulary, comprehension and science is reported on a class by class basis; in all but one classroom there were significant changes on most measures, with medium to large effect sizes. When classes were analysed in aggregate, the magnitude of improvements on pre- to post-test measures were: *d*=0.78 for	(i) This study would have benefited from more sophisticated statistical analyses of nested/grouped data from across the two schools and five classrooms (linear modelling)

	qualitative data was gathered to measure fidelity of implementation.		science, *d*=0.51 for vocabulary and *d*=0.33 for comprehension. The researchers concluded that variation in how integration was enacted accounted for variation in results from classroom to classroom, e.g. the non-significant improvement in scores in one classroom was attributed to the teachers' low confidence levels in science.	
Feldwisch (2014) United States	This research examined the enactment of the Arts Integration Program (AIP) in 11 schools. This programme was developed by an unnamed national non-profit, who solicited the researchers/authors to examine the implementation of the programme over a three-year period. Pre-post data was gathered using a range of tools: observations, interviews, pre/post literacy assessment (designed for the programme being reviewed).	Single Group N=11 schools Performance on literacy assessment gathered from *n*=43 students (Year 2) and *n*=190 students (Year 3)	Students were described as being highly engaged in arts integrated lessons, demonstrated in eye contact, avoidance of off-topic talk, behavioural indications of excitement (e.g. smiles); teachers indicated that students were highly enthusiastic about the lessons; during artist residencies, student engagement was even more pronounced. Test scores improved from the start to the end of the year in each of the years 2009/2010 (Year 2) and 2010/2011 (Year 3), however these differences were only significant in the 2009/2010 year.	(i) Limited information on the nature of field notes taking during observations (ii) Limitations in the nature of the quantitative data collected - not all students provided consent (potential for sampling bias), variation in numbers from year to year;

			From the student perspective, a number of challenges were reported, which varied by grade level: e.g. students grew frustrated when asked to revise their work; the literacy assessment was complex and time-consuming.	
Frankel (2015) United States	This quasi-experimental study draws on a large-scale data set (gathered over a three-year period) to examine the impact of a writing-visual integrated programme, implemented 2-3 times a week by teachers in the middle elementary grades. 'Picture writing/image-making' involves students in hands-on artwork which feeds into their subsequent writing. The programme addresses both art and literacy state standards. Pre-Post data was gathered from: art and writing samples scored using an instrument from previous research in this area; state reading and writing assessments; Gates-MacGinitie Reading Comprehension test; scores district-level writing assessments)	Between Groups *N*=6 schools *N*=1500 students (approx.) Treatment: *n*=3 schools Control: *n*=3 schools	In each year of the study, students in the treatment condition made significantly larger gains in both writing/visual literacy scores when compared to the control group (large differences are noted, but no effect sizes are reported); gains were seen across demographics (e.g. SEN, EAL, gender). A higher proportion of treatment group students reached proficiency in the grade 5 NECAP writing test. Students scored significantly higher in most aspects of the district writing prompt assessment (e.g. voice), when compared to comparison schools. Slight gains were seen in the state reading results (NECAP), with gains in on only year of the Gates-MacGinitie Reading test data.	(i) Some test items scored by teachers, others by independent researchers. (ii) Distinctions are drawn between some treatment schools, with better results seen in one 'high fidelity' school. (iii) There was high staff turnover during the course of the study (iv) Lack of qualitative data to unpack the processes/challenges

Graham (2016) United States	This quasi-experimental study examined the impact of STEAM lessons (emerging as a result of different forms of professional development for teachers) on physical science learning in grades 3 to 5 using performance on standardised district tests of science as a measure.	Between Groups *N*=7838 (5th Grade) Treatment: *n*=2156 Control: *n*=5682;	The study found that students exposed to the STEAM lessons demonstrated greater improvement on physical science benchmark assessments than students exposed to a STEM-only physical science curriculum. With models controlling for confounders, Cohort 1 (teachers only after PD) saw moderate improvement in scores compared with control students (0.35SD) and Cohort 2 (teachers + teaching artists) had slight improvements in benchmark scores over the control group (0.10SD). The authors made the following statement to demonstrate the impact of their results: *this amounts to an improvement, with a student moving from 50th percentile to 63rd percentile in the targeted curriculum when assigned a teacher well-trained in the STEAM curriculum, all other factors equal.*	(i) Intact groups (ii) Limited range of outcome measures (i.e. only Science was assessed), (iii) Potential novel effect (9 hours)
Gray (2022) United States	This multi-site cluster-randomised control trial study examined the effectiveness of 'Zoology One: Kindergarten Research Labs', developed by American Reading Company. This is a full-year curriculum that involves 120	Between Groups *N*=1589 (Kindergarten)	Treatment group scored higher on: passage comprehension; (Glass's Delta ES=0.16); letter naming fluency (Glass's Delta ES=0.28);	(i) A researcher-constructed item was used for science (ii) A very large block of time (90 minutes) was available of integrated literacy/science

	minutes of daily integrated science/literacy instruction (e.g. read alouds, direct instruction in reading/writing, levelled texts, hands-on science inquiry). Researchers randomly allocated classrooms to the treatment/control condition within the schools. A number of literacy and science outcome measures were examined in this study.	*N*=71 classrooms across 21 schools. Every school had at least one classroom in the treatment and control conditions.	reading motivation (Glass's Delta ES=0.32). There were no difference on measures of: word attack, word identification, word comprehension, science knowledge, general reading outcome (DRA), writing outcomes. These effects were replicated within subgroups (language, gender etc). Fidelity of implementation had varying effects on different measures.	instruction daily - this may not be possible/replicable in other jurisdictions
Hardiman (2019) United States	This study examined the effects of arts-integrated lessons on long-term memory for science content. The paper describes the results of a randomised control trial involving 5th grade learners that measured retention of science content using arts-integrated science units and matched units employing convention science instruction. This study involved classroom-matched pairs and equivalent control groups with condition reversal. Pre-, Post- and Delayed Post-test data was gathered using researcher-designed curriculum-based assessments.	Between Groups *N*=350 (5^{th} Grade)	No statistically significant difference between percent of retained content in the arts-integrated instruction condition versus the conventional instruction condition was noted. 'Basic' readers remembered significantly more science content learned through the arts at the delayed post-test than basic readers who learned science through conventional methods.	(i) Possible novel effect (arts-integrated science lessons were taught over a period of 4-6 weeks) (ii) Arts-Integration was a pedagogy in this study - art content was not a key element of the intervention (iii) Researcher-designed outcomes measures (based on the content taught) (iv) Possible experimenter effect (classroom teachers taught both the arts-integrated and conventional curricula)

			There was no effect of arts-integrated instruction for different science units (e.g. Astronomy, Chemistry, Life Science etc.) by student reading level. A treatment by order effect was found: students who took arts-integrated science in the first session remembered more science in the second session when they learned science through conventional lessons.	
Harris (2019) United States	This study examined the impact of the integrated/interdisciplinary approach to teaching on 4th grade learners' achievement in standardised tests in literacy, numeracy and science. Test scores (PARCC [Literacy/Numeracy], ASK-4 [Science]) from 2014-2017 for a randomly selected grouping of 50 schools using integrated–interdisciplinary curriculum and 50 schools using subject-specific curriculum were analysed to address this research question.	Ex-Post Facto Design *N*=100 schools (4th Grade) Treatment: *n*=50 schools Control: *n*=50 schools	A significant difference was found to indicate that an integrated–interdisciplinary curriculum increased 4th grade students' academic achievement in language arts as measured by the New Jersey PARCC. In all 3 years of the study, the curriculum was the strongest predictor of language arts student achievement scores. However, other variables were also significant predictors (e.g. ELL, SES) for different years. A significant difference was found to indicate that an integrated–interdisciplinary curriculum increased 4th grade	(i) Limited observations (ii) Very specific focus on use of resources (iii) No collection of student data; some data only provided by sub-sample of teachers. (iv) Lack of comparison group

students' academic achievement in mathematics as measured by the New Jersey PARCC. In all 3 years of the study, the curriculum was the strongest predictor of mathematics student achievement scores. However, other variables were also significant predictors (e.g. ELL, SES) for different years.

A slight significant difference was found to indicate that an integrated–interdisciplinary curriculum increased fourth grade students' academic achievement in science as measured by the New Jersey ASK4-Science assessment. In 2015, the strongest predictor of science achievement was negatively impacted by the SES predictor, but in 2016 and 2017, the curriculum was the strongest predictor of science student achievement scores.

The influence of the integrated–interdisciplinary curriculum over time shows a slight increase in scores for language arts and mathematics, but a slight decrease in science scores.

Hraste (2018) Croatia	This study examined the efficacy of an integrated mathematics/geometry and physical activity (PA) program for 4th grade pupils. Pre-post data was gathered using a researcher-designed assessment of relevant geometry concepts as well as a National School Program approved mathematics test.	Between Groups *N*=36 (4th Grade) Treatment: *n*=19 Control: *n*=17	The results of factorial mixed design between/within the 2×2 ANOVA showed a statistically significant impact of 'group' on achievement i.e. participants who gained their mathematics and geometry knowledge through the integrated Maths/PA approach were significantly more successful than the control group.	(i) Intact groups (ii) Researcher designed materials (iii) No underlying discussion informing the integration of PA and maths (i.e., seems 'surface level') (iv) PA and PE are not interchangeable - PA is a part of PE (i.e. generalisability issues) (v) Short intervention time/novel effect. (vi) Small sample size
Inoa (2014) United States	This multi-stage cluster randomized control trial study examined the impact of arts (theatre) integrated instruction on state standardised test scores in mathematics and ELA/literacy. Eight schools were involved in this study four schools involved in the control and treatment condition.	Between Groups *N*=1193 (6th/7th Grade) Treatment: *n*=464 Control: *n*=729	At a 6th grade level, students in the arts integration group were significantly more likely to reach proficiency in state mathematics assessments, but not in language arts; there was a significant difference in mean scores for literacy and mathematics. At a 7th grade level, there was no significant difference in the number of students who reached proficiency in language arts/maths; neither was there a significant difference in mean scores for language arts/maths.	(i) Results had to be disaggregated by grade level (6/7) due to a lack of homogeneity when pooled in the same sample - the authors point out that this reduced statistical power. (ii) No substantial/theoretical reason is offered for why scores were greater in the 6th grade sample than the 7th grade sample. (iii) No direct measures of *arts* knowledge/skills are reported - maths/literacy assessments may not be the most appropriate measures.
Jia (2021) United States	This experimental research examined the impact of a novel unit of work that aligned with the interdisciplinary principles of STEAM education and 'Maker' education on student's learning motivation, self-efficacy, and	Single Group *N*=164 (3rd Grade)	Learning Motivation: The mean values for the dimensions of total score, attention, relevance, and satisfaction were all >3 (where 5 is the highest score) indicating	(i) Role of research team is somewhat unclear in the write up (possible researcher effects) (ii) Insufficient information on some of the instruments used (validity/reliability issues) (iii) Write-up of quantitative results lacks clarity

	acquisition of interdisciplinary knowledge. Measures included: Learning motivation scale (adapted from Keller, 2009); modified version of General Self-Efficacy Scale (GSES); STEAM Test questions (from a multidisciplinary test bank); a researcher designed. The learning activities were designed using the engineering design framework as a guide.		(according to the authors) that learners were motivated to complete STEAM tasks. Self-Efficacy: Students' levels of self-efficacy to complete integrated STEAM tasks had significantly increased by the end of the intervention. STEAM Scores: Students' acquisition of relevant STEAM knowledge following completion of the course was positive with an average score of 65.46 calculated from the assessment data gathered.	(iv) Hawthorne effect
Lamb (2015) United States	This quasi-experimental study examined the content, cognitive, and affective outcomes related to STEM integrated curriculum for elementary school learners. The intervention used in this study was a whole school STEM integration curriculum designed by the STEM coordinator and district office. Pre-post data was gathered across a number of measures: Science Interest Survey (from Lamb et al., 2012), Self-Efficacy in Science and Technology Short Form (from Lamb et al., 2014), Paper Folding Task (Ekstrom et al. 1976) and the Shepard Metzler Test of Mental Rotation Task (Shepherd & Metzler, 1971),	Between Groups *N*=254 (Kindergarten, 2nd Grade, 5th Grade) Treatment: *n*=111 Control: *n*=143	An estimate of effect size of the mean group difference across the statistically significant constructs revealed that performance was practically better among those students in the Treatment condition: self-efficacy (*d*=1.27), science interest (*d*=1.97), spatial visualization (*d*=1.30), and mental rotation (*d*=1.42). There were no real differences between the groups on content learned. The ANOVA F-test results	(i) Intact groups (ii) Suitability of some outcome measures (e.g. cognitive tests) (iii) Insufficient attention to potential maturation effects in the analysis (but was present in the discussion) (iv) Sampling biases (only certain age groups involved) (v) Some self-report measures used (prone to bias)

	Science & Mathematics Content Exam (Pearson)		show that there are statistically significant differences on the measured constructs across the main effect of one group, accounting for 13% of the variance in scores between the groups (large). Post hoc tests indicated that many of the differences between the control and comparison groups happened in the 2nd or 5th grade, thus indicating that access to STEM education at earlier ages can have 'pay-off' at later stages.	
LaMotte (2018) United States	Students in an quasi-experimental group took part in a unit on transportation that integrated dance movements. The control group studied the same content, but without dance integration. Researcher-designed tests on transportation and dance concepts involving multiple-choice/short answer/matching/true-false questions along with journal entries were used to measure the outcomes of this study.	Between Groups *N*=40 (5th Grade) Treatment: *n*=18 Control: *n*=22	There was a significant increase in knowledge of transportation scores in the experimental group; this was not the case for the control group. There was a significant difference in post-test transportation scores between the experimental and control group. There was no increase in scores on the movement test from pre to post (this test was only taken by the experimental group). The experimental group produced more affective responses in journals than the control group (49 affective v 10 affective). 80% of	(i) Intact groups (ii) Statistical analysis does not account for multiple comparisons - other statistical tests may have been more appropriate (iii) Some measures only taken by the experimental group (iv) Non-conventional reporting of inferential test results. (v) Small sample size.

			affective responses were positive in the experimental group. The experimental group also produced more psychomotor responses (53 v 0). Some negative responses to dance-integration were reported.	
Luna (2015) United States	In the context of the school garden, two second-grade teachers introduced children to homonyms, linking science and literacy. This study examines the impact of this teaching approach on student homonym knowledge. Pre-post and extended post-intervention data was gathered using a researcher-designed measure of student homonym knowledge and application.	Single Group *N*=43 (2nd Grade; 2 classrooms)	Statistically significant improvement in homonym knowledge in the short term (2 weeks) and long term (7 months).	(i) No comparison/control group (ii) Small sample size (iii) Researcher-designed measure. (iii) Very particular instantiation of integration. (iv) Not clear that it was the *garden* context that made a specific contribution to the growth in homonym knowledge
Luo (2022) United States	This study examined elementary students' understanding of variables having received instruction using a maths-computational thinking integrated approach. Cognitive think-alouds were used as an indicator of learning for a small subset of the participants (*n*=9) alongside responses to a researcher-designed assessment.	Single Group *N*=36 (4th Grade)	Analysis of both the quantitative and qualitative data gathered indicated that students lacked a conceptual understanding of using variables to create generalized problem solutions that could work with any set of inputs. Students also had difficulty with specific mechanics of using variables e.g. updating variable values.	(i) High attrition from mid- to post-intervention (ii) Unclear who delivered instruction to the students (i.e. researchers or teachers) (iii) Analysis of quantitative data was short and was inadequately described. (iv) Insufficient discussions on the maths-CT integrated lessons
Makopoulou (2020)	This study relies on the premise that integrating dance (as a form of physical activity) into reading lessons will support	Between Groups	Quantitative data suggest that reading comprehension improved in the dance-integrated group;	(i) Intact groups (ii) Use of intact rather than randomised groups

United Kingdom	children's reading comprehension (based on performance of a test of reading comprehension with unknown provenance) and other affective outcomes. Students in the intervention group took part in dance-integrated lessons that involved dance-based responses to text. Pre-post intervention data was gathered.	*N*=42 (Year 4 students) Treatment: *n*=24 Control: *n*=18	linear mixed model results indicated a moderate to large effect size (ES=.61). Qualitative data indicated that overall, students were able to pinpoint aspects of new learning from the intervention. Not all students agreed that the dance-integration benefited their reading ability. Most (80%) students enjoyed the dance-integrated lessons.	(iii) Lack of detail on the measures used (reading comprehension in particular) (iv) Short duration of the intervention (8 lessons) (v) the control group took part in regular PE lessons and thus had less exposure to text
Miller (2019) Australia	This experimental study examined how mathematical knowledge and thinking, specifically the identification of mathematical patterns and structures can be promoted through engagement with coding lessons involving a visual programming language ('Scratch'). Pre-post data was gathered using a researcher designed assessment. Observations of participants involved in the treatment group were also conducted.	Between Groups *N*=135 (Year 2) Treatment: *n*=40 Control: *n*=95	Both the control and treatment groups exhibited a significant increase between pre- and post-test scores with regards to their understanding of mathematical patterning. However, the intervention group performed much better (M: 16:30) than the control group (M: 11.95). Analysis of the qualitative data indicated that the students involved in the intervention group developed a strong conceptual understanding of pattern and were able to apply it to other, related mathematical concepts e.g. algebra.	(i) Researcher delivered the intervention (ii) Inequivalent sample sizes for control and treatment group (iii) Activities of the control group were somewhat unclear.
O'Neal (2017)	This study used the 2011 Trends in International Math and Science Study (TIMSS)	Ex-Post Facto Design	Student achievement in life science/biology was correlated	(i) Comparisons occurred at the education system level

Multiple Contexts	data to examine inter-subject correlations (biology, maths, science) on student achievement for 4th and 8th grade learners. The study aimed to investigate if TIMSS data could be utilised to establish guidelines on STEM integration so as to assist practitioners and direct the course of future research.	*N*=58 Educational Systems (4th Grade) *N*=50 Educational Systems (8th Grade)	with achievement in mathematics and other sciences across three analytical areas: mathematics and science student performance, achievement in cognitive domains, and achievement in content domains. The importance of linking student learning experiences between and within STEM domains to support high performance on TIMSS assessments was indicated by correlations of moderate strength (57<r<85) between life science/biology and mathematics content domains, as well as by stronger correlations (73<r<97) between life science/biology and other science domains. According to the author, results demonstrated the foundational nature of STEM knowledge at the fourth grade level, and established the importance of strong interconnections among life science/biology, mathematics, and other science. The results from this investigation promote a holistic design of	(ii) Not all STEM subjects were considered in the analysis (iii) Cross-country comparisons are inherently problematic due to differences in national curricula (or indeed their absence) (iv) Standard measurement issues when working with International Large Scale Assessments

			school learning opportunities to improve student achievement in life science/biology and other science, technology, engineering, and mathematics (STEM) subjects at the elementary and middle school levels	
Panagopulos (2015) United States	Two distinct studies were involved in this dissertation. Study 1 was a longitudinal, quasi-experimental study that examined the impact of arts-integrated instruction on the reading scores of students qualifying for the free and reduced meal benefit (FARMS) at five different schools compared to similar peers in schools that do not use arts-integrated instruction. Three years of standardised test data in reading contributed to this study. Study 2 examined the arts-integration practices of teachers using classroom observations and semi structured interviews.	Between Groups *N*=727 (Grade 3) Treatment: *n*=344 Control: *n*=383	Study 1: Within all five arts-integrated schools, the mean change for FARMS students exceeded mean change for non-FARMS students. There were no statistically significant differences in reading change scores between non-FARMS and FARMS students. Mean reading change scores for the Grade 3 cohort of students showed that students in AI schools outperformed students in non- arts-integrated schools during the year that they were in Grade 3 (2011). This trend did not continue through grades four and five (but the small effect sizes calculated indicated that those in arts-integrated schools did perform somewhat better that those in a non-arts-integrated setting).	(i) Generalisability difficulties (different forms of art integration may be used in different schools due to inconsistencies in definitions). (ii) Suitability of outcome measures and Maryland formula to calculate growth (iii) Non-integrated AI classrooms were not observed (making some statements difficult to have complete confidence in). (iv) Contextual factors may have influenced teacher responses/practices (curricular reform in Maryland). (v) Sample sizes were inconsistent (see p. 60)
Peppler (2014)	This paper examines the impact of a school arts program (Learning and Achieving Through the Arts - LATA) involving a	Between Groups *N*=6 schools	Chi-square analysis of the proportions of students achieving or surpassing a 'proficient' score	(i) More sophisticated statistical analysis needed to fully account/control for differences between

United States	community arts organisation and a school district. Control schools taught the arts as part of their curriculum but did not engage in arts integration using the LATA model. Standardised ELA test scores from state assessments were compared across treatment and control schools.	Treatment: *n*=3 schools Control: *n*=3 schools	across treatment/control schools indicated that: baseline scores were higher in the control schools but from Year 1 of implementation they were consistently higher in the treatment schools. Analysis of ELLs' scores showed even more substantial differences between conditions.	schools (chi-square analysis is somewhat limited). (ii) analysis based on school-level data rather than data for individual students. (iii) other confounding variables (relating to inter-school differences) not accounted for (iv) ELA scores may not be the best measure (how is learning in the arts captured?)
Robinson (2021) United States	This study examined how an integrated STEM teaching model influenced 5th grade students' perceptions of their mathematics and engineering abilities. The teaching model used a 'real life' issue (access to clean water) as a learning context. Pre-, mid- and post-intervention surveys that used items from the following instruments: (1) The Patterns of Adaptive Learning Survey, (2) Mathematical Attitude Assessment, (3) Engineering Skills Self-Efficacy Scale, and (4) Intersectionality of Non-Normative Identities in the Cultures of Engineering Survey	Single Group *N*=17 (5th Grade)	Quantitative data indicated a decrease in mathematics self-efficacy but an improvement in perceived mathematics usefulness from mid-unit to post-unit. There was no statistically significant change in students' Engineering Self Efficacy scores throughout the intervention. The qualitative data indicated an increase in students' confidence to do difficult math at the end of the unit. However, previous experiences with maths appeared to influence students' overall progress towards higher levels of self-efficacy for mathematics. The authors asserted that integrated teaching can foster 'positive shifts'	(i) No role for technology in the study (ii) Unequal emphasis on disciplines (iii) researcher bias may have been present due to their involvement in the unit's delivery (iv) Sample size (v) No specific achievement data

			in students' perceived STEM abilities.	
Sáez-López (2016) Spain	This quasi-experimental study analysed the potential benefits that coding with a visual programming language ('Scratch') may have on the affective and academic functioning of learners when it is integrated with the science- and art-based subjects. The study occurred over a two-year period where students participated in 20 one-hour coding sessions that were integrated with science and arts concepts. Pre-Post data intervention gathered using Visual Blocks Creative Computing Test [VBCCT]. Post-intervention data was gathered using researcher-designed questionnaire that addressed topics related to active learning, knowledge of art history concepts, computational concepts, perceived usefulness of course, and enjoyment. Structured observations were also conducted.	Single Group *N*=107 (6th/6th Grade)	Significant improvements in Visual Blocks Creative Computing Test [VBCCT] from pre- to post-intervention; Students achieved an above-average understanding of art and history concepts; High levels of perceived usefulness; Working with visual languages provided fun, motivation, enthusiasm, and commitment from the student.	(i) Intact groups (ii) Inadequate description of intervention - the roles of teachers and learners etc. in this study are very unclear (iii) Use of VBCCT to infer improvements in computational thinking skill mays have been flawed given the absence of a control group
Samuels (2019) United States	This piece of action research was spurred by the researcher's observation that some students in their class, including children from a low SES background, were uninterested in social studies and were disengaged/disruptive as a result. The treatment group (taught by the researcher) in this action research undertook a thematic unit on World War One over the course of ten lessons (two weeks); history was integrated with art, math, science, music, ELA and literature/drama. Data gathered included a	Between Groups *N*=40 (7th Grade) Treatment: *n*=26 Control: *n*=14	Analysis (descriptive) of student interest surveys indicated that treatment group students were no more or less likely to signal social studies as a preferred subject after the intervention (only pre-data were collected with the control group); on direct measures of attitudes towards social studies, attitudes declined post-intervention.	(i) Intact groups (ii) Irregularities in statistical analysis i.e. multiple t-tests used instead of ANOVA. (iii) Short duration of study (iv) Potential researcher effects (researcher worked with control group; another teacher with the treatment) (v) Researcher indicates that students were uneasy/ hesitant in responding during interviews (v) Limited statistical analysis (small quant data set) on some data, use of multiple t-tests rather than ANOVA (or another analysis)

	researcher designed survey of student subject preferences; pre/post-test of multiple-choice questions on WW1 knowledge (drawn from a state provided test bank); researcher-designed Likert questionnaire of attitudes to social studies; observations; semi-structured interviews.		Analysis of student scores on the WW1 knowledge test indicated that both the treatment and control group made statistically significant gains from pre to post, but that there was no difference between each group at post Limited analysis of student responses to interviews indicated that they were positive about the unit, but others stated "it was all over the place and very confusing"; there were mixed opinions on student perceptions of thematic. *vs* 'traditional' teaching; some evidence that students who did poorly on the social studies test were more likely to express negative sentiments towards thematic teaching. Observational data indicated that student interest in the unit decreased as time passed;	(vi) Confounding variables likely had an impact on the results reported
Santaolalla (2020) Spain	This study examines how pre-service teachers' perceptions of interdisciplinary teaching were improved through designing and enacting a unit of work in a local National Archaeological Museum (Study 1). These units were subsequently enacted with primary school children (Study 2). Mixed results are	Between Groups *N*=58 (7th Grade) Treatment: *n*=33 Control: *n*=25	Based on ANOVA analyses of the researcher-constructed items, all students (both control and test groups) improved their knowledge; while those in the treatment group demonstrated improved scores, it was not possible to conclude that	(i) Quantitative analysis completed with very small sample and group sizes that were non-equivalent (Study 1) (ii) Researcher-designed instrument; ambiguity in reporting of some results

	reported. Findings for Study 2 are summarised here. This was a quasi-experimental study that gathered pre- and post-intervention data from students on their perceptions of the museum as a place for learning, their knowledge of social science (e.g. Prehistoric times, Roman Hispania) and knowledge of mathematics (e.g. symmetry, polygons). Students in the treatment group worked with pre-service teachers to complete interdisciplinary activities while in the museum while those in the control group completed a traditional tour of the museum.		interdisciplinary learning at the museum explained the difference in learning.	
Saraniero (2014) United States	This 3-year quasi-experimental study compared the effects on student learning of two contrasting approaches to teacher professional development in arts integration – a summer institute model and a model combining the summer institute with instructional coaching. The study was organised by the DREAM (Developing Reading Education with Arts Methods) initiative (based in the state of California).Learner performance was evaluated using the California Standards Test Language Arts test. Demographic data from teachers along with evaluations of their lesson plans (as evaluated by project-based rubrics).	Between Groups *N*=116 (In-Service Teachers) Treatment 1 (Institute Only): *n*=25 teachers per annum Treatment 2 (Institute and Coaching): *n*=25 teachers per annum Control: *n*=39 teachers per annum (no	In-Service Teachers (N=116) Coached teachers reported greater confidence integrating the arts, produced higher-quality work samples, taught more reading concepts with arts integration, implemented more arts standards, and used arts integration more frequently than did the institute-only teachers or the control group teachers. There were significant differences found between teachers' students in the 3rd Grade pre-test but no significant difference between groups in third grade post-test.	(i) Significant funding attached to the study (e.g. teachers received stipends to participate in the PD course/study) (ii) Sample size (iii) Post-Hoc tests seemed to indicate that on some measures there were no statistically significant differences between the two treatment groups, even though descriptive statistics showed higher scores for the institute+coached group (e.g. p. 11). Non-use of other post-hoc tests makes it difficult to interpret what type of professional development supported improved student outcomes. (iv) Content of the PD delivered was not clearly explained (v) Student test-scores were not equivalent in the pre-test stage (making interpretation more complex)

		Professional Development)	There were significant differences found between the coaching and comparison groups in the 4th Grade pre-test and post-test. However, a lack of post-hoc tests on the pre-test data makes it difficult to completely interpret these findings. Overall, findings were mixed.	(vi) Inadequate consideration for potential confounding variables.
Schugar (2017) United States	This study involves a secondary analysis of how US NAEP (National Assessment of Educational Progress) reading comprehension scores are impacted by particular factors. One of the factors examined was 'cross-curricular reading', which researchers constructed from individual items in the student survey that accompanies the comprehension items. This component/factor was built from responses to questions on the frequency of reading texts such as paperbacks, soft cover books, puzzle books or magazines for science, social studies or history.	Ex-Post Facto Design *N*=165000 (4th Grade; Academic Year 2004/2005)	Based on the within school model, every increase of 1 standard deviation in the frequency with which students engaged in cross-curricular reading was matched with a 1.10 point increase in informational text comprehension; this effect was not associated with FARMS status. Based on the between school model, a 1 standard deviation increase in the frequency with which cross-curricular reading materials were used was associated with an increase of 3.46 points in comprehension. The authors conclude that the use of cross-curricular reading materials (reading across the curriculum) are associated with	(i) NAEP relies on student report of practices. (ii) the 'cross-curricular reading' variable relies on measures that do not specifically measure this construct

			increase in reading comprehension	
Smith (2016) United States	Students in the experimental group engaged in dance-integrated geography/history (anthropology) lessons on Ancient Egypt. Students in the control group experienced 'business as usual' teaching. For example, they enacted - through dance- the geographical features of the greater Nile region. Pre-post data was gathered using the textbook chapter text and an adapted attitudinal survey on enjoyment of social studies. Field notes were also completed after each lesson	Between Groups *N*=56 (6th Grade) Treatment: *n*=28 Control: *n*=28	At post-test, students in the intervention group scored higher on the chapter test; there was no difference in attitude to social studies between the groups. Analysis of journals indicated that students expressed positive views towards dance (notably, journals were not conducted with the regular instruction group) but there were also negative sentiments "It's kind of hard for me to learn anything this way" (Female. White.)."	(i) Intact groups (ii) Though measures of knowledge were reported as being the same in both groups at time 1, a larger study would be needed for greater causal claims. (iii) Researcher effects are also very likely given the researcher enacted the dance unit
Smith-Gayle (2014) United States	This study involves a secondary analysis of state testing data on ELA scores, to examine the impact of integrated teaching on middle school boys' performance.	Ex-Post Facto Design *N*=3448 middle school boys (6th, 7th and 8th Grades from 2007-2012) Treatment: *n*=2318 Control: *n*=1130	There was no significant difference in academic achievement between boys who attended schools that used an integrated or a traditional curriculum. Analysis of sub-groups indicated that there were no significant differences at individual grade levels based on traditional *vs* integrated instruction.	(i) Potential confounding variables could not be accounted for. (ii) Categorical scores used in the ELA test examined may have not been an appropriate measure of progress. (iii) Intact groups (iv) Modelling does not take full account of other variables nested in classes/schools - more advanced modelling procedures needed to address the full complexities of the data. (v) Limited information provided on the nature of the integrated teaching (and how it differed to 'traditional' teaching) in selected schools
Snyder (2014)	Using a quasi-experimental approach, this study examined the impact of the Supporting	Between Groups	For all grade levels, the overall percentage of students who	(i) Intact groups

United States	Arts Integrated Learning for Student Success (SAILSS) model on student achievement three years after its introduction to a struggling 'Tier I' school in the United States. Data gathered included: State and local standardised testing, School-Level Environment Questionnaire [SLEQ], Arts Integration: Classroom Observations for Middle Schools [AICOM], arts integration logs and researcher-designed parent, student, and teacher surveys.	*N*= 2 schools (*n*=1171, Grades 6-8) Treatment: *n*=1 school, *n*=510 students approx. Control: *n*=1 school, *n*=661 students	scored Proficient or Advanced from the treatment school was statistically higher than the percentage of students from the control school to receive the same grades (p<0.01). Regarding Reading, the effect size representing the improvement in standardised test scores in the Treatment school from 2009 to 2012 for Grades 6, 7 and 8 respectively was *d*=0.29, *d*=0.53 and *d*=0.24. For standardised test scores in mathematics, the effect sizes calculated for the improvements in test scores from 2009 to 2012 were *d*=0.41, *d*=0.48 and *d*=0.07 for Grades 6, 7 and 8 respectively. An ANOVA found that time did not have any impact on In addition to increasing student achievement on state-wide assessments, implementing this arts integration model positively correlated with a 77% decline in discipline referrals, and overall positive change in school climate based on teacher, staff, student, and parent perception.	(ii) Sample issues (generalisability of the experiences of specialist middle school teachers to generalist teachers)

Swan (2013) United States	This quasi-experimental study investigated the 'Preparation for Future Learning' (PFL) approach, which has similarities to problem-based learning. In this approach, students encounter a scenario or problem before receiving more formal instruction on how to address the problem. More specifically, students encountered the problem relating to world water shortages and allocations in social studies, before learning about proportional allocation in mathematics and subsequently applying this knowledge in science (water cycle problems) and language arts (persuasive writing and oral presentations which communicate new learning). Thus integrated learning happens in a consecutive rather than concurrent manner. Students receiving PFL instruction were compared to those who did not. Outcomes were measured using researcher-designed data literacy assessment, written student reflections and teacher interviews. Pre-post data was gathered.	Between Groups *N*=576 (7th Grade) Treatment: *n*=114 Control: *n*=462	Students in the intervention group outperformed students in the non-intervention group. Analysis of individual items in the pre/post assessment revealed varying improvements from pre- to post-intervention. Students reported benefits of the sequenced, integrated teaching in qualitative data; teachers reported benefits for student learning.	(i) Uneven numbers in intervention and non-intervention groups (ii) The data is nested within classes and schools, therefore more advanced modelling/statistical analyses may have better accounted for the data (including differences between schools) (iii) There is a lack of detail on the provenance of the unit (iv) Limited information on the activities completed by the control group and the fairness of the comparison
Talbert (2019)	Two studies are involved in this dissertation which explored the relationship of background knowledge, reading comprehension, and content learning on student progress and learning. Study 1 involves an experimental design that examines the effects of teaching inferential strategies while building knowledge using informational text. Pre-post intervention data	*Study 1* Between Groups *N*=94 (5th Grade) Treatment 1 (Inference Instruction): *n*=32	Study 1: Both intervention groups performed better on a researcher-created measure of reading comprehension when compared to a business-as-usual control, though not on a standardized measure of reading comprehension (Gates MacGinitie Reading Comprehension, *p*=.544).	Study 1 (i) Learners in the study were all 'average' or 'above average' readers (sampling/ generalisability biases) (ii) Relatively small sample size (iii) Intervention length was quite short (thus making it unlikely that the standardised test scores would ever improve in the study)

	was gathered using a standardised measure of reading comprehension (Gates MacGinitie Reading Comprehension) as well as a researcher-designed measure of reading comprehension and disciplinary content.. Study 2 is a meta-analysis that aims to determine if the practice of integrating science and literacy instruction is associated with higher effect sizes for measures of literacy and science achievement.	Treatment 2 (Content Knowledge Instruction): *n*=32 Control: *n*=32 *Study 2* Meta-analysis (*N*=32 studies)	There were no statistically significant differences between the content knowledge and inferential intervention groups on a measure of content learned, indicating either method of strategy instruction was effective for knowledge acquisition. The effect sizes calculated indicate that those in the inferential intervention group may have had a slight advantage on performance on content knowledge group. Study 2: Results from 32 studies show an overall weighted mean effect of 1.04 for science outcomes and .245 for literacy outcomes - evidence that the practice of integrating science and literacy instruction is effective.	(iv) Researchers (rather than teachers) delivered the entirety of the intervention. Study 2 (i) Publication bias (ii) Number/type of studies included
Tank (2014) United States	This quasi-experimental study examined the integration of STEM teaching and learning in elementary classrooms with the reading of nonfiction children's literature. This doctoral research investigated how this particular approach to integration impacted student learning in each of these disciplines. Particular attention was paid to the impact of	Between Groups *N*=120 (5th Grade) *Study 1* Treatment (Science+ Reading): *n*=27	Study 1: The addition of the integrated nonfiction science reading unit had a significant effect on pre and post-test science content assessment performance (*d*=0.28). The author noted that in their analysis of the interview data that the treatment students were	(i) Intact groups (ii) Validity of measures used (iii) Gaps in the baseline data gathered means that some interpretations may not be fully appropriate (iv) Explanation of work undertaken in the treatment/control conditions were unclear.

science and engineering integrated activities on student outcomes which involved: Landforms Unit Pre-Post Assessment; 'Stick in the Mud' Unit Assessment; Teacher-developed measures of Reading Comprehension and Vocabulary Assessment; Student Interviews; Document Analysis [Student notebooks]).	Control (Science): *n*=27 *Study 2* Treatment 1 (Science+ Engineering): *n*=26 Treatment 2: Science+ Engineering+ Reading: *n*=24	more frequently able to recall science concepts in their interviews than the control students. While both the control and treatment group improved on their post-test performance, the results from the statistical analysis of the reading content assessment found a non-significant effect of the treatment on students' learning in reading for those students in the science condition. Study 2: Treatment students who participated in the integrated nonfiction unit had increased learning in engineering as seen by their higher average notebook scores, especially for the open-ended notebook tasks, and by their ability to more frequently recall and make connections between engineering concepts in their student interviews. The treatment students from the engineering+science+reading condition did significantly outperform the control students on the reading content assessment.

Tucker (2017)	This action research examined the effects of arts-based instruction on student literacy achievement for 5th grade learners. Students received an arts-based literacy intervention (delivered by the researcher) three times a week for number of weeks. Pre-Post data from a researcher-designed assessment of literacy skills and a researcher-designed survey of students' attitudes about reading/literacy were gathered.	Single Group *N*=10 (5th Grade)	An increase in student achievement was noted in the researcher-designed assessment of literacy skills i.e. 8 out of the 10 students showed some improvement between pre- and post-assessment.	(i) Researcher bias (researcher administered the intervention to a group of children known to them) (ii) Selection bias (students were selected for the intervention because they were reading below expected grade levels - any extra intervention regardless of form may have caused an improvement) (iii) Insufficient control for potential confounders (iv) Inadequate description of intervention
Vallera (2015) United States	Students in the treatment group of this quasi-experimental study participated in a ten-day agriculture/STEM integrated curriculum, including a trip to an agricultural education centre. This integrated curriculum included challenge-based, technology-integrated curriculum and was enacted by the regular classroom teacher (but developed by the researcher). The control group did not receive agriculture-related instruction. Post-intervention outcome measures included: knowledge test of science/technology/engineering and agricultural literacy [KnowASTE, based on combination of items from other standard measures]; attitude measure on science/technology/engineering and agricultural literacy [ThinkASTE, based on combination of items from other standard measures]; analysis of project-based performance tasks.	Between Groups *N*=95 (4th Grade) Treatment: *n*=42 Control: *n*=38	ANOVA revealed significant difference between groups (partial η^2 = .192) and from pre to post-test (partial η^2 = .187) favouring the treatment group. ANCOVA analysis controlling for pre-test scores maintained this difference (partial η^2 = .336). Analysis of sub-scale scores of science/technology/engineering and agriculture scores, controlling for pre-test scores also revealed significant differences between groups. In relation to attitudes/beliefs, a significant difference between groups was also found in ANOVA analysis (partial η^2 = .253) with a difference from pre to post-test	(i) Intact groups (ii) Treatment group did not receive similar content instruction - potentially unfair comparison;

			also evident (partial η^2 = .082). ANCOVA analysis again revealed a difference while controlling for pre-test scores (partial η^2 = .188).	
van't Hooft (2012) United States	This article provides an overview of the middle school 'Thinking with Data' (TWD) curriculum which aims to guide teachers in their efforts to teach data literacy across multiple subject areas. The study compared the impact this interdisciplinary curriculum had on students' data literacy skills compared to those who received a more traditional curriculum. Pre-Post data gathered included: researcher designed data literacy assessments, subject-specific classroom test scores (science, mathematics).	Between Groups *N*=576 (7th Grade) Treatment: *n*=114 Control: *n*=462	Students in the treatment condition had a gain score that was 3 points higher on a 15-point test of data literacy than comparable students who did not engage with the TWD materials. This difference was statistically significant with a very large effect size, *d*=1.24. MANOVA analysis indicated that approximately 31% of the difference in total gain scores between the conditions can be attributed to the 'Thinking With Data' curriculum. Learning gains in discipline specific science and maths assessments were noted.	(i) Imbalanced sample sizes across conditions (ii) Generalisability of school context i.e. four teachers taught data literacy integrated with their individual subject (rather than all subjects being taught together by one teacher) (iii) Pre-test scores indicated that the students involved in the study may not have been directly comparable (iv) Imprecise statistical analysis which casts doubt over the results obtained e.g. the assumptions for homogeneity of variance and covariance were violated (V) Test instruments were not treating consistently in terms of administration or analysis
Volk (2017) Slovenia	This quasi-experimental study examined how the use of tablet devices supported the integration of maths with learning science and the Slovene language, drawing on concrete, visual and abstract representations. The topics of time and orientation (maths) served as the basis for the integration. Post-intervention data was gathered using the following materials: researcher-designed	Between Groups *N*=12 schools *N*=259 (Average age of learners was 8 years) Treatment: *n*=124 (across 6 schools)	Students in the treatment group scored significantly higher at the second taxonomy level (procedural knowledge; *r*=0.33) and third taxonomy level (problem solving; *r*=0.30). There was no significant difference at the first taxonomic level (conceptual knowledge), which was predicted by the	(i) Intact groups (ii) Key contextual factors likely had an impact on the unit's success e.g. learners had a high level of musical knowledge thanks to their prior educational experiences (iii) Range of data gathered

	assessment addressing a three-level taxonomy - conceptual knowledge, procedural knowledge, problem-solving; observation; field notes.	Control: *n*=135 (across 6 schools)	authors. Qualitative data indicated that students found tablets more engaging to learn with.	
White (2014) United States	This study examined the effect of having 3rd-8th grade teachers participate in professional development course on integrating literacy practices into science teaching. The study relies on teacher self-report of practices as well as a variety of student measures, including state achievement tests.	Single Group *N*=39 (In-Service Teachers)	Student achievement data showed some positive gains (e.g. 81 more students reached proficiency in informational text in the year the project took place; slight gains in proficiency in the Earth and Space Science portion of the state achievement test), however, these results were not uniform or particularly obvious.	(i) Limited information provided on the precise measures/methodology used (ii) Reliance on self-report of practice (iii) State science scores only available for two grades (5th/8th) (iv) No comparison/control group (v) State assessment data examined in aggregate only - not individual student improvements
Wright (2017) United States	This quasi-experimental study tested the effects of teaching literacy skills to kindergarteners in the context of science inquiry units (SOLID Start Curriculum), drawing on disciplinary literacy/language. Learners in the first treatment group completed the Weather then Plant unit. Learners in Treatment 2 completed the Plant then Weather unit. The control group had 'business as usual' teaching. Pre- and post-test data from students were gathered using the following: Peabody Picture Vocabulary Test; Expressive Vocabulary test; SOLID start interview (a researcher designed interview protocol to assess scientific reasoning and vocabulary).	Between Groups *N*=147 (Kindergarten) Treatment 1 (Weather/Plant Unit): *n*=61 Treatment 2 (Plant/Weather Unit): *n*=41 Control: *n*=45	Children in the treatment conditions outperformed those in the control condition on all four parts of the SOLID Start interview (claim; evidence-based support, receptive vocabulary; use of science vocabulary in context; $p < .001$ for each; Hedges's $g > 0.7$ for each). Multiple linear regression of various predictors indicated that the intervention had the largest effect on post-test scores when science knowledge and oral language were held constant.	(i) Limited information on the BAU group (ii) High levels of scaffolding/support provided to teachers (iii) intervention took place over a limited time frame (iv) limited amendments made to units at local level
Zhang (2012)	This quasi-experimental study examined the effectiveness of the Integrated	Between Groups	After controlling for gender, SES, pre-test scores on attitude	(i) Intact groups

China	Experiential Learning Curriculum (IELC) in China. This curriculum was developed to engage Chinese elementary students in science in relation to a range of other disciplines to cultivate a scientifically literate society by focusing science instruction on practical applications of scientific knowledge. Pre- and post-intervention data was gathered using the following measures: Researcher design measure of student attitude about science, researcher designed measure of student citizenship beliefs. Researcher-designed post-measures of student attitude toward the learning environment were also gathered.	*N*=385 (Elementary Students) Treatment: *n*=201 Control: *n*=184	towards science, the authors found that IELC treatment positively affected student attitude toward science, despite the influences of the student background information and their previous attitude. Consequently, the authors The IELC improved students' attitude toward science, their citizen beliefs, and their attitudes about the learning environment. Significant improvements around citizen beliefs were only noted for 3rd grade learners.	(ii) Short intervention period (1 year)/potential novelty effect (iii) Significant level of PD provided to treatment group (during the summer and regularly throughout the year from the researchers) (iv) Groups were not comparable in terms of age/SES (v) Suitability of research measure (vi) Limited range of data gathered i.e. no learning outcome data

Section 2

Methods Used to Source Research on Pedagogical Approaches

Section 2

Methods Used to Source Research on Pedagogical Approaches

This appendix outlines the manner in which relevant empirical and conceptual literature were sourced to inform the material on *pedagogical approaches* reviewed in this report. As outlined in Chapter 1 of the Report, a bespoke search and analytic approach was required as conducting a full systematic or scoping review on the broad concept of pedagogy would be an impractical task.

Research Questions

This review sought to answer the following research questions:

- What pedagogical approaches and considerations should inform a redeveloped primary school curriculum?
- How can these be enacted in an integrated context?

Identifying Relevant Studies

The researchers intended to provide a wide-angle, research-informed view of pedagogy (and integration), but it was beyond the scope and time limitations of the project to provide a systematic mapping of all relevant literature. Consequently, four strands or methods were used to identify relevant literature to answer the research questions. They involved:

1. Teaching Effectiveness Studies
2. Scoping Searches
3. Handsearch of seminal texts on pedagogy / pedagogical approaches
4. Content analysis of studies on curriculum integration

The approaches used to identify the above sources will now be described.

Teaching Effectiveness Studies

A range of pertinent studies, reports and papers that focussed on effectiveness were outlined through a search of academic databases and grey literature sources such as Google Scholar. Forward and backward chaining were used to identify linked studies/reports. For example, analysis of Coe et al.'s (2020) review of studies led to the

identification of further such reviews that were considered appropriate for inclusion. The papers included in the synthesis are outlined in Table 1. Readers should note that this table does *not* purport to be a comprehensive or exhaustive catalogue of papers in this area. However, the very high level of similarity across the papers is noteworthy.

Table 1 Headline findings from reviews of the research on effective teaching

Citation	(Coe et al., 2020a)	(Coe et al., 2014)	(Ko et al., 2013)	(Husbands & Pearce, 2012)
Characteristics of good teaching	Four overarching dimensions of effective teaching: 1. Understanding the content and how it is learnt 2. Creating a supportive environment for learning 3. Managing the classroom to maximise learning opportunities 4. Presenting content, activities and interactions that activate student thinking	Six components of 'great' teaching: 1. (Pedagogical) Content Knowledge 2. Quality of instruction 3. Classroom climate 4. Classroom management 5. Teacher beliefs 6. Professional behaviours	Effective teachers are: 1. Clear about instructional goals 2. Knowledgeable about curriculum content and strategies for teaching it 3. Communicate clear expectations to students 4. Make expert use of instructional materials 5. Use knowledge of students to adapt instruction and anticipate misconceptions 6. Teach metacognitive strategies 7. Address both higher and lower-order cognitive objectives 8. Monitor student understanding and provide feedback 9. Integrate teaching within/across subjects 10. Accept responsibility for student outcomes	Effective pedagogies: 1. Give serious consideration to student voice 2. Depend on teacher behaviour, knowledge, understanding and beliefs 3. Involve clear thinking about both short-term and long-term goals 4. Build on pupils' prior learning and experience 5. Scaffold pupils' learning 6. Involve a range of techniques and configurations (e.g. guided/structured; whole-class/group/individual) 7. Develop higher-order thinking and meta-cognition; drawing on dialogue/questioning 8. Embed assessment for learning 9. Are inclusive, taking learners needs and equity into account
Methodological approach	'Umbrella' review of existing reviews combined with a database search	Expert review of existing research evidence (methodology unclear)	Expert review of existing research evidence (methodology unclear)	Expert review of existing research evidence (methodology unclear)

Citation	(Kyriakides et al., 2013); (Creemers & Kyriakidēs, 2012)	(Siraj et al., 2014)	(Muijs et al., 2014)	(Seidel & Shavelson, 2007)
Characteristics of good teaching	The following characteristics were found to make a significant (but not necessarily large) contribution to learning: 1. Orientation* 2. Structuring* 3. Questioning* 4. Teacher Modelling* 5. Application* 6. Classroom as learning environment* 7. Time management* 8. Assessment* 9. Self-regulation 10. Concept-mapping 11. Computer use 12. Interpersonal behaviour 13. Classroom organisation * = *factors included in the dynamic model of educational effectiveness*	Effective pedagogical strategies included: 1. Organisation 2. Shared goals/objectives 3. Homework 4. Classroom climate 5. Behaviour management 6. Collaborative learning 7. Personalised learning 8. Making links explicit 9. Dialogic teaching and learning 10. Assessment for learning 11. Plenary	Behaviours associated with student achievement include a focus on: 1. Opportunity to learn and time on task 2. Instruction and interaction 3. Classroom climate 4. Teacher expectations 5. Self-regulated learning 6. Non-cognitive outcomes (e.g. motivation)	Framework (based on Bolhuis, 2003) includes the following factors that can be used to evaluate teaching: 1. Knowledge domain 2. Time for learning 3. Organisation of learning 4. Social context 5. Goal setting/orientation 6. Execution of learning activities 7. Evaluation 8. Regulation and monitoring
Methodological approach	Meta-analysis of 167 studies on how teachers behaviours impact student outcomes; informed by the dynamic model of educational effectiveness (Creemers & Kyriakidēs, 2012)	Longitudinal study of 3,000 students in the UK (Effective Pre-School, Primary and Secondary Education, 2-16); strategies above noted from observations of teaching	Expert review of existing research evidence (methodology unclear)	Meta-analysis of effectiveness studies

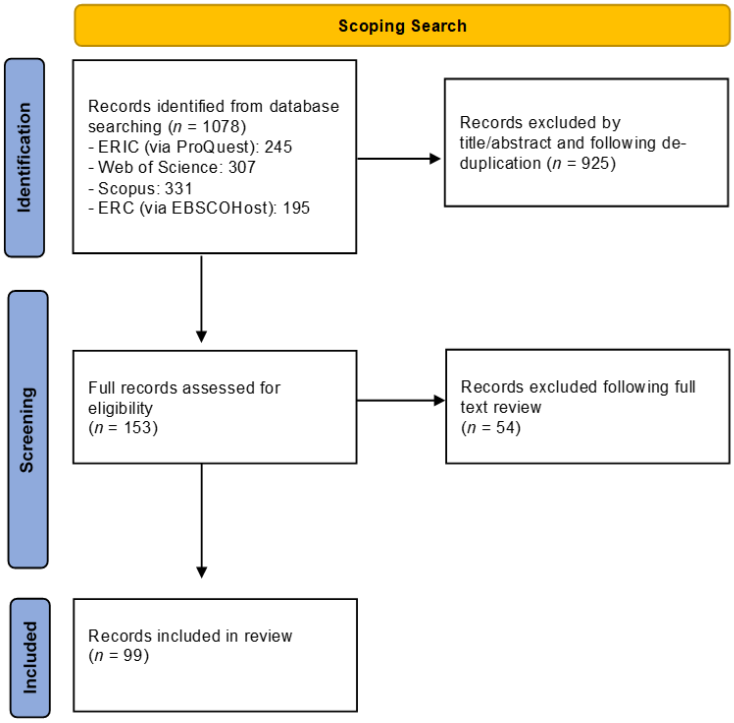
Scoping searches

A Boolean search strategy was developed and deployed in relevant databases (Web of Science, Education Resources Information Center (ERIC), Education Research Complete (ERC) and Scopus) to conduct a 'review of reviews' relevant to pedagogical approaches. Search terms were informed by seminal research within the field, pilot searches and consultation with DCU's librarians. Specifically, the stems "pedagog*", "instruct*" or "teach*" were paired with phrases that restricted the return of studies to those that involved systematic reviews, meta-analyses or broad reviews of best evidence relevant to primary school-aged learners. Given the breadth of literature that could have been returned, results were restricted to peer-reviewed articles written in English published over the past decade. Inclusion and exclusion criteria are outlined in Table 2. A summary of the search strategy and search strings used for each database is provided later in this section. Studies brought forward for inclusion are included in column 2 of Table 3.

Table 2 Eligibility Criteria for Database Searches

Category	Inclusion Criteria	Exclusion Criteria	Rationale
Population	Study included the equivalent of Irish primary school aged children (4-12 years)	Study did not include equivalent of Irish primary school-aged children.	To conduct a broad search that focused on the population of interest
Language	Written in English or Gaeilge	Written in any other language other than English or Gaeilge	Reviewers are fluent in English and Gaeilge
Time Period	2012 to present	Any publication outside of these dates	To reflect the most recent literature and research on assessment
Study Focus	Publications examining pedagogical approaches used in primary educational contexts	Publications that do not discuss pedagogical approaches used in primary educational contexts.	To build a broad evidence base on what can be considered good assessment practice in primary educational contexts
Source Type	Systematic Reviews, Meta-Analyses, Scoping Reviews (peer-reviewed)	Papers reporting single studies	

Figure 1 PRISMA Flow Chart Summarising Inclusion of Review Studies on Pedagogical Approaches



Handsearch of seminal texts on pedagogy / pedagogical approaches

Studies returned through the Boolean searches outlined above were supplemented by targeted hand searching of evidence sources (e.g. Education Endowment Foundation, What Works Clearinghouse), handbooks (e.g. The Sage Handbook of Curriculum, Pedagogy and Assessment) and journal articles based on individual studies or conceptual pieces. This included further searches of tools such as Google Scholar. Extensive forward and backward chaining was also used to identify pertinent review/empirical/efficacy studies relevant to a given pedagogical approach. Studies identified using these strategies are included in column 3 of Table 3.

Content analysis of studies on curriculum integration (Report 1 Annex)

In order to illustrate the pedagogical approaches adopted in the context of integrated teaching, all 211 studies captured in the systematic review from Report 1 were re-analysed. Where the pedagogical approaches were clearly outlined, they were tagged and added to column 4 of Table 3. A certain level of caution is required in interpreting this column; in many instances, the pedagogical approaches were not clearly described in an individual study. Therefore, the studies cited are best treated as illustrative or

demonstrative rather than conclusive in terms of the breadth, proportionality or frequency with which certain pedagogical approaches are used in integrated contexts.

Limitations

It is important to note a number of limitations associated with the methodological approach adopted. Though the reliance on multiple strands of evidence provides a broad base on which assertions can be made, it is not possible to conclude that every possible or relevant piece of empirical evidence has been considered. The research team aimed to mitigate against the omission of important evidence through backward/forward chaining and the critical consultation with five peer-reviewers drawn from disparate disciplinary backgrounds. Nonetheless, gaps and omissions cannot be ruled out. The reliance on review studies (e.g. meta-analyses) ensures that only approaches that have had extended consideration in the academic literature are included; less commonly used approaches are, therefore, not included. The reliance on review studies is, also, not beyond reproach. Meta-analyses, in particular, require cautious interpretation. Furthermore, pedagogical approaches that may demonstrate promise, but that have not been tested quantitatively, would not have been included in such reviews

Table 3 Pedagogical Approaches: Summary of Key Sources Consulted

	Scoping Search *Boolean searches of academic databases to identify research reviews.*	**Handsearch** *Handsearches included review of seminal books, Google Scholar searches, backward and forward chaining of key papers.*	**Content Analysis of Report 1 Annex** *Re-analysis of all 211 studies to identify examples of the approach as applied in an integrated context.*
Collaborative Approaches *Approaches involving some level of interaction between students (e.g. in pairs, groups, whole-class)*	Lai, 2022; Pelligrini, 2021; Savelsbergh, 2016; Tenenbaum, 2020; Arizmendi, 2021; Dietrichson 2021	Cohen, 1994; Johnson, 2009; Kaendler, 2015; Kyndt, 2013; Leung, 2015; van Leeuwen, 2019; Webb, 2009; Webb et al., 2009	Atalay, 2015; Bartels, 2019; Hieu, 2019; Fazio, 2019; Edsall, 2012; Luna, 2015; Lovemore, 2021; An, 2014; An, 2013; An, 2017, An, Tillman, 2014; Cassidy, 2022; Evans, 2015; Odegaard, 2014; Mildenhall, 2021; Birchinall, 2013; van't Hooft, 2012; Schellinger, 2021; Cannon-Ruffo, 2020; Dogan, 2019; Convertini, 2020; Ensign, 2012; García-Carrillo, 2021; Huck 2019; Kim, 2015; Lau, 2018; Vallera, 2015; Feldwisch, 2014; Zhang, 2012; Lehrer, 2021; Hourigan, 2021
Explicit teaching, direct teaching & gradual release of responsibility *Approaches that emphasise overt modelling and explanation in advance of student application of new learning*	Filderman, 2022; Graham, 2012; Gillespie & Graham, 2014; Bogaerds-Hazenberg, 2021; Stevens, 2019; Peterson, 2020; Monei & Pedro, 2017; Powell, 2021; Schnepel, 2022; Clausen, 2021; Okkinga, 2018; Roesslein, 2019; Kang, 2015; Arizmendi, 2021; Gersten, 2020; Kaldenberg, 2015; Stewart, 2020; Rodgers, 2022; Pyle, 2017	Adams, 1996; Engelmann, 1980; Engelmann, 1988; Kirschner, 2022; Pearson, 1983; Pearson, 2019; Rosenshine, 1986; Rosenshine, 2008; Stockard, 2018; Tobias, 2009; Ashman, 2021; Archer, 2011	Cervetti, 2012; Bazemore, 2015; Hawley, 2022; Robinson, 2021; Odegaard, 2014; Jia, 2021; Viñas, 2021; Tytler, 2021; Calder, 2013; Talbert, 2019; Bravo, 2014; Revelle, 2019; Duke, 2021
Play-based learning *Use of playful approaches to support learning*	Jerebine, 2022a; 2022b; Burson, 2022; Guan, 2022	Baker, 2021; Bergen, 2013; Burghardt, 2010; Doebel, 2023; Eberle, 2014; EEF, 2023; French, 2022; Gray, 2013; Guan, 2022; Hirsh-Pasek, 2009; Hirsh-Pasek, 2015; Jensen, 2019; Johnstone, 2022; Knight, 2016; Lillard, 2013; Liu, 2017; Mardell, 2023; O'Keeffe, 2023; Parker, 2019; Parker, 2022; Plass, 2015; Plass, 2020; Pyle, 2017; Rogers, 2010; Siraj-Blatchford, 2010; Siraj-Blatchford, 2002; Skene, 2022; Sylva, 2004; Weisberg, 2013; Weisberg, 2016; Wood, 2014; Zosh, 2018;	Wright, 2017; Monteiro, 2021; Edwards, 2016; Collins, 2016; Tam, 2021; Speldewinde, 2022
Project-, problem- and inquiry-based learning *Broadly related approaches*	Koyunlu Ünlü, 2022; Savelsbergh, 2016; Slavin, 2012; Estrella, 2018	Alfieri, 2011; Brush, 2017; Chen, 2019; Estrella, 2018; Ferrero, 2021; Friesen, 2013; Furtak, 2012; Krajcik, 2022; Lazonder, 2016; Mergendoller, 2018; Walker, 2009	*Project:*Atalay, 2015; Fazio, 2019; Jordan, 2016; Nadelson, 2014; Baptiste, 2022; van't Hooft, 2012; Bungum, 2014; Sáez-López, 2016; Öztürk Yılmaztekin, 2017; Revelle, 2019; Speldewinde, 2022; Rule, 2021; Huck, 2014; Trent, 2018; Khanna, 2021; Harris, 2019; Havice, 2018; McDowall, 2019; Trinter, 2021; Miller-Ray, 2019; Savage, 2016; Vallera, 2015; Duke, 2021; Aranda, 2020; Ollila, 2019 *Problem:* Quigley, 2019; Hieu, 2019; Edsall, 2012; Lovemore,

			2021; An, 2013; An, 2017; Wieselmann, 2021; Robinson, 2021; Cassidy, 2022; Evans, 2015; Baker, 2017; Birchinall, 2013; Birsa, 2018; Viñas, 2021; Tytler, 2021; Schellinger, 2021; Convertini, 2020; Ensign, 2012; García-Carrillo, 2021; Gomez Zwiep, 2016; Miller, 2019; Kloser, 2017; Lamb, 2015; Havice, 2018; Kok, 2014; Rico, 2020; Zhang, 2012; Hourigan, 2021 *Inquiry:* Aguirre-Munoz; 2021; Quigley, 2019; Cervetti, 2012; Batic, 2020; Gray, 2022; Jordan, 2016; Liston, 2018; An, 2017; Robinson, 2021; Nesmith, 2017; Odegaard, 2014; Evans, 2015; Nadelson, 2014; Wright, 2017; Baptiste, 2022; Birchinall, 2013; Björklund, 2017; Tytler, 2021; Calder, 2013; Bravo, 2014; López-Leivaa, 2016; Cannon-Ruffo, 2020; Collins, 2016; Convertini, 2020; White, 2014; Kim, 2015; Khanna, 2021; Hardiman, 2019; Kok, 2014; Leszczynski, 2014; Levy, 2018; McDowall, 2019; Graham, 2016; Savage, 2016; Jamil, 2017; Gerke, 2017; Lehrer, 2021
Scaffolding *Various forms of scaffolds including visuals, concrete materials, graphic organisers*	Filderman, 2022; Peterson, 2020; Powell, 2021; Kang, 2015; Roesslein, 2019; Arizmendi, 2021; Kim, 2022; Kul , 2018	Van De Pol, 2010; Vygotsky, 1978; Wood, 1976	Aguirre-Munoz; 2021; Lovemore, 2021; An, 2017; Mildenhall, 2021; Cannon-Ruffo, 2020; Dogan, 2019; Cotič, 2021; Ensign, 2012; Volk, 2017; Atalay, 2015
Technology enhanced approaches *Including device use, software, virtual/augmented reality, flipped learning, machine/AI learning, robotics, makerspaces, 3D printing, game-based learning*	Chauhan, 2017; Stringer, 2022; Martínez-Soto; 2023; Mikropoulos 2022; Boon, 2020; Deunk, 2018; Pelligrini, 2021; Benavides-Varela, 2020; Cheung, 2013; Savelsbergh, 2016; Slavin, 2012; Archer , 2014; Zheng , 2016; Verschaffel, 2019; Vasquez, 2016; Rodríguez-Jiménez, 2023; Wang, 2023; Sokolowski, 2015; Wen, 2022; Van Schoors , 2021; Fotaris, 2017; Cai, 2022; Santhanasamy, 2022; Li, 2022; Sanusi, 2022; Papadopoulos, 2020; Zhong, 2018; Xia, 2018; Zhang, 2021; Zhang, 2022; Benitti, 2012; Hein, 2018; Lee, 2023; Rouse, 2022 Chang , 2022; Guan , 2022; Pellas, 2019; Sun, 2020; Lei, 2022; Hussein, 2019; Wang, 2022; Akça, 2021	Celik, 2022; Chiu, 2023; Lewin, 2019	Quigley, 2019; Fazio, 2019; Mildenhall, 2021; Öztürk Yılmaztekin, 2017; Cannon-Ruffo, 2020; Harris, 2015; Volk, 2017; Huck 2019; Leszczynski, 2014; Rico, 2020; Ensign, 2012; García-Carrillo, 2021; Rosenthal, 2020; Mildenhall, 2021; Tam, 2021; Miller, 2019
Supports for Individual Needs *Includes specific interventions approaches for learners with individual needs.*	*The following studies returned in the database search focused on a range of individual needs (e.g. reading difficulties, emotional and behavioural needs). They were reviewed as part of the analysis and cited as appropriate amongst the other pedagogical approaches (Chapter 4) or in chapter 3 (e.g. Deunk et al., 2018 provides a review of differentiation; this was cited in chapter 3 under inclusive pedagogy).* Graham, 2021; Neitzel, 2022; Stentiford, 2018; Peterson, 2020; McKenna, 2021; Monei, 2017; Powell , 2021; Schnepel, 2022; Clausen, 2021; Puzio, 2020; Deunk, 2018; Pellegrini, 2021; Benavides-Varela, 2020; Lein, 2020; Cheung, 2013; Wanzek, 2013; Barbier , 2022; Roesslein, 2019; Flynn , 2012; Kang, 2015; Pico, 2021; Arizmendi, 2021; Ralston, 2014; Wanzek, 2016; Gersten, 2020; Vasquez, 2016; Kaldenberg, 2015; Dietrichson, 2021; Swanson, 2017; Goldfield, 2022; Dobinson, 2021; Gist, 2022		

Scoping Searches: Pedagogy

Details of Database Searches

Database: Scopus via Elsevier
Date: 4/3/23
Searcher: Patrick Burke

#	Searches	Results
1	TITLE-ABS-KEY (pedagog* OR instruct* OR teach*)	1,461,154
2	S1 AND TITLE-ABS-KEY ("primary school" OR "primary edu*" OR "elementary school" OR "elementary edu*" OR "early child* edu" OR "early years edu*" OR "middle edu*" OR "middle school")	58,485
3	S2 AND TITLE-ABS-KEY ("systematic review" OR "meta-analysis" OR "best evidence"))	413
4	S3 AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012)) AND (LIMIT-TO (LANGUAGE , "English"))	331

Database: Web of Science via Clarivate
Date: 11/3/22
Searcher: Patrick Burke

#	Searches	Results
1	TI=(pedagog* OR instruct* OR teach*)	394,852
2	AB=(pedagog* OR instruct* OR teach*)	723,529
3	Author Key Words [AK] =(pedagog* OR instruct* OR teach*)	178,347
4	S1 OR S2 OR S3	935,480
5	S4 AND ALL=("primary school" OR "primary edu*" OR "elementary school" OR "elementary edu*" OR "early child* edu" OR "early years edu*" OR "middle edu*" OR "middle school")	39,287
6	S5 AND ((TI=("systematic review" OR "meta-analysis" OR "best evidence")) OR AB=("systematic review" OR "meta-analysis" OR "best evidence") OR AK=("systematic review" OR "meta-analysis" OR "best evidence"))	372
7	S6 AND Limiters applied: 01-01-2023 to 11-03-2022; English Language	307

Database: ERIC via ProQuest
Date: 11/3/22
Searcher: Patrick Burke

#	Searches	Results
1	title(pedagog* OR instruct* OR teach*)	276,603
2	abstract(pedagog* OR instruct* OR teach*)	595,989
3	Keyword - if(pedagog* OR instruct* OR teach*)	98,013
4	S1 OR S2 OR S3	696,141
5	S4 AND "primary school" OR "primary edu*" OR "elementary school" OR "elementary edu*" OR "early child* edu" OR "early years edu*" OR "middle edu*" OR "middle school"	90,044
6	S5 AND (title("systematic review" OR "meta-analysis" OR "best evidence") OR abstract("systematic review" OR "meta-analysis" OR "best evidence") OR if("systematic review" OR "meta-analysis" OR "best evidence"))	347
7	S6 AND Limiters applied: English language; 01-01-2012 to present	245

Database: Education Research Complete (ERC) (via EBSCOHost)
Date: 11/3/23
Searcher: Patrick Burke

#	Searches	Results
1	TI pedagog* OR instruct* OR teach*	281,524
2	AB pedagog* OR instruct* OR teach*	787,205
3	KW pedagog* OR instruct* OR teach*	113,796
4	S1 OR S2 OR S3	837,676
5	S4 AND "primary school" OR "primary edu*" OR "elementary school" OR "elementary edu*" OR "early child* edu" OR "early years edu*" OR "middle edu*" OR "middle school"	75,891
6	S5 AND (TI ("systematic review" OR "meta-analysis" OR "best evidence") OR AB ("systematic review" OR "meta-analysis" OR "best evidence") OR KW ("systematic review" OR "meta-analysis" OR "best evidence"))	257
7	S6 AND Limiters applied: English language; 01-01-2012 to present	195

Section 3

Methods Used to Source Research on Assessment Approaches

Section 3
Methods Used to Source Research on Assessment Approaches

This section describes how the relevant empirical and theoretical literature that informed the *assessment approaches* discussed in Chapter 6 for primary-school-aged learners were identified. Given the potential breadth and depth of literature involved in reviewing such a topic, a bespoke search approach was undertaken.

Research Questions

This scoping review was guided by the following research questions:

1. What approaches and considerations should inform a redeveloped primary school curriculum?
2. How can these be enacted in an integrated manner?

Identifying Relevant Studies

Three broad sources of information were used to identify relevant literature to answer the above research questions:

1. Seminal Texts/Handbooks
2. Scoping and Hand Searches
3. Content Analysis of Report 1 Annex

The identification of literature relevant to each category involved a bespoke search strategy. These will now be described.

Seminal Texts/Handbooks

Using the authors' knowledge of the field of assessment, seminal texts and handbooks along with documents relevant to the Irish research context were identified. For example, handbooks authored by leaders within the field of classroom assessment (e.g. Brookhart, McMillan, Wiliam) were first identified, examined and included pending their relevance to Irish primary classrooms. Work by Irish authors working within the field of educational and classroom (e.g. Murchan & Shiel, 2017) were also consulted. Key Irish policy documents relevant to primary education were also considered to be highly pertinent e.g. NCCA, 2007; Lysaght et al., 2019. For certain assessment approaches, highly

relevant literature known to the authors was consulted e.g. standardised tests (O'Leary et al., 2019), questioning (Rosenshine, 2012). While knowledge of the field is advantageous to the identification of core texts and ideas, it is prone to bias. Consequently, other sources of information were also drawn upon.

Scoping and Hand Searches

In February and March 2023, a Boolean search strategy was developed and deployed in relevant databases (Web of Science, Education Resources Information Center (ERIC), Education Research Complete (ERC) and Scopus) to conduct a 'review of reviews' of different assessment approaches. Search terms were informed by seminal research within the field, pilot searches and consultation with DCU's librarians. They included "assessment", "classroom assessment", "formative assessment", "summative assessment", "assessment for learning", "assessment of learning". These terms were paired with phrases that restricted the return of studies to those that involved systematic reviews, meta-analyses or broad reviews of best evidence relevant to primary school-aged learners. Given the potential breadth of literature that could have been returned on assessment, the identification of overview studies like these seemed most prudent given the timeframe of this research project. Results were restricted to peer-reviewed articles written in English published over the past decade. A summary of the search strategy and search strings used for each database is available at the end of this section.

The second search strategy involved hand searching techniques. This included a manual search of two relevant journals (*Assessment in Education: Principles, Policy and Practice*, *Irish Educational Studies*) to identify any single studies or theoretical papers that would support the review. Recent evidence reviews, such as those conducted by the Education Endowment Foundation (EEF) were also examined for relevance. The reference lists of all included sources of evidence were screened for additional potentially relevant studies.

Study Selection

For the scoping review, a total of 1376 relevant studies were identified through database searching. In the initial evaluation of the selected literature, the second author read the title and abstract of each study to determine its relevance in accordance with the eligibility criteria outlined in Table 4. This resulted in a total of 72 studies requiring full text review. Both researchers independently read the same 5 papers to agree on the application of the eligibility criteria. Disagreements were resolved through discussions

related to the inclusion and exclusion criteria.

Table 4 Eligibility Criteria for Database Searches on Assessment

Category	Inclusion Criteria	Exclusion Criteria	Rationale
Population	Study included the equivalent of Irish primary school aged children (4-12 years)	Study did not include equivalent of Irish primary school-aged children.	To conduct a broad search that focused on the population of interest
Language	Written in English or Gaeilge	Written in any other language other than English or Gaeilge	Reviewers are fluent in English and Gaeilge
Time Period	2012 to present	Any publication outside of these dates	To reflect the most recent literature and research on assessment
Study Focus	Publications examining the assessment practices and approaches used in primary educational contexts	Publications that do not discuss assessment practices and approaches used in primary educational contexts.	To build a broad evidence base on what can be considered good assessment practice in primary educational contexts
Source Type	Systematic Reviews, Meta-Analyses, Scoping Reviews (peer-reviewed)	Papers reporting single studies	

An additional 22 papers (highly relevant single studies, additional meta-analyses) were subsequently identified through the hand searching process. A total of 47 works were included in the final corpus of the scoping review which included peer-reviewed journal articles and reports. A summary of the article search and review process can be found in Figure 2.

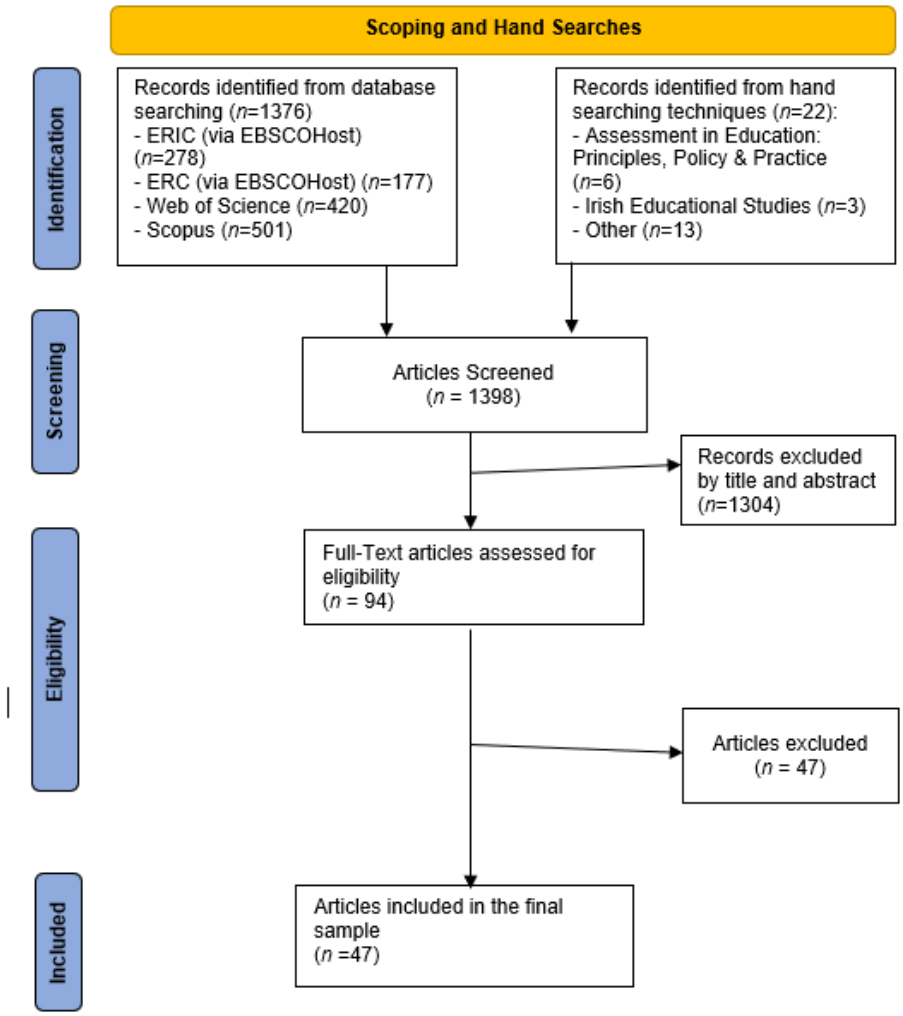


Figure 2 Outline of the article search and selection process

Content Analysis of Report 1 Annex

To demonstrate what assessment approaches were adopted in the context of integrated teaching, the 211 studies contained within the systematic review from Report 1 were re-analysed. Where the assessment approaches were clearly outlined, they were 'tagged' and recorded. It should be noted that a level of caution should be applied to the process undertaken here. For most of the studies included in Report 1's annex, the role of assessment received little attention and was inadequately described from a theoretical or practical perspective. Due to insufficient information, the analysis of the included studies' assessment practices by the research team may be incomplete and may not fully reflect their intended or actual use. Therefore, the categorisation of the assessment approaches

used in these studies should be seen as *illustrations* as to how certain assessment approaches are used in integrated contexts

Analysis and Reporting

A narrative synthesis was undertaken using a textual, rather than a statistical, approach to draw conclusions on the assessment practices that should be considered in a redeveloped primary curriculum. This synthesis drew on previously discussed a priori assessment concepts and practices (e.g. formative assessment strategies; Wiliam & Leahy, 2015) and inductively-developed categories of practices. The sources (and their origins) are described in Table 5. These were then described qualitatively in Chapter 6 of the current report.

Limitations

The research team notes that certain limitations need to be considered when interpreting the findings of the studies reviewed. For example, only English-language articles published in the last decade were included in the scoping review, which inevitably limited the depth and breadth of the review. Additional, relevant articles may have been unintentionally omitted due to the databases searched and/or the search terms used. Furthermore, many of the studies in the scoping searches examined the use of assessment methods across a range of educational settings. Their investigation of assessment methods included studies that involved learners in primary, secondary and tertiary contexts. Given the qualitative differences in teaching primary-aged learners, some of the findings from these studies must be carefully interpreted to determine their relevance and applicability to primary classrooms. Finally, risk of publication bias or a quality appraisal of studies was not conducted given the heterogeneity of studies. However, relevant details on the studies have been included where possible to give an indication of the quality of research conducted.

Table 5 Assessment Approaches: Summary of Sources

Assessment Methods	Seminal Texts/ Handbooks *Handsearches included review of seminal books, highly relevant empirical/theoretical literature and Irish policy documents*	Scoping and Hand Searches *Boolean searches of academic databases to identify research reviews and the supplementary hand searches arising from the literature returned from the scoping searches (see Figure 2)*	Content Analysis of Report 1 Annex *Re-analysis of all 211 studies to identify examples of the approach as applied in an integrated context.*
Classroom Tests These were isolated, classroom based activities that learners had to respond to within a designated period of time e.g. curriculum-based tests, drawing assessments, chapter tests, teacher-designed tests, researcher-designed tests, questionnaires/surveys, multiple choice questions, drawing etc.	Brookhart (2020); Chen (2021); Hattie (2019); Lysaght (2019); McMillan (2020); Murchan (2017); NCCA (2007); Scully (2017)	Alqassab (2023); Brookhart (2016); Chen (2022); Chang (2020); Collins (2018); Cutumisu (2019); Graham (2015); Harris (2022) Hartmeyer (2018); Heitink (2016); Killian (2021); Klute (2017); Lane (2020); Long (2022); Pyle (2017); Tang (2020); Wellberg (2023); Wheadon (2020); Zhang (2019); Zheng (2020)	Akbar (2012); Alghamadi (2017); Atalay (2015); Bazemore (2015); Birsa (2018); Bravo (2014); Brugar (2012); Casady (2015); Cannon-Ruffo (2020); Cervetti (2012); Duke (2021); Fazio (2019); Fragakis (2019); Gray (2022); Halimah (2021); Hardiman (2019); Hraste (2018); LaMotte (2018); Levy (2018); Liston (2018); Luna (2015); Luo (2022); Mård (2022); McDowall (2019); Miller (2019); Sáez-López (2016); Samuels (2019); Schellinger (2021); Smith (2016); Snyder (2014); Talbert (2019); Tank (2014); Trent (2018); Tucker (2017); Tytler (2021); van't Hooft (2012); Vallera (2016); White (2014)
Feedback Any information (e.g. written, oral) regarding a learner's progress towards a particular objective that comes from any outside agent including peers and teachers (written or oral)[1]	Brookhart (2020); Cizek (2019); Davidson (2010); Hattie (2007); Hattie (2019); Lysaght (2019); McMillan (2020); Murchan (2017); Newman (2021); NCCA (2007); William (2015)	Alqassab (2023); Baliram (2018); Bartholomew (2019); Coe (2019); Double (2019); Education Endowment Foundation (EEF) (2021b); Graham (2015); Harris (2022); Hartmeyer (2018); Heitink (2016); Hodgen (2018); Hodgen (2020); Karaman (2021); Klute (2017); Killian (2021); Koenka (2021); Lane (2019); Lee (2020); Li (2020); Merrit (2022); Pandero (2013); Schildkamp (2020); Wisniewski et al. (2020); Zhang (2019)	Aguirre-Munoz (2021); Cunnington (2014); Garcia-Carillo (2021); Kirsten (2019); Maitles (2012); Mård (2022); McDowall (2019); Mildenhall (2021); Moss (2019); Quigley (2019); Trent (2018); Tytler (2021)
Observations This involved the collection of learner progress on an informal, ongoing basis during learning experiences e.g. field notes (as noted by teachers only)	Brookhart (2020); Cizek (2019); Lysaght (2019); McMillan (2020); Murchan (2017); NCCA (2007); William (2015)	Alqassab (2023); Chang (2020); Harris (2022); Miller-Bains (2017); Pyle (2017); Zhang (2019)	Bazemore (2015); Brough (2012); Feldwisch (2014); Follong (2022); Khanna (2021); Lau (2018); Luna (2015); McDowall (2019); Panagopulos (2015); Sáez-López (2016); Samuels (2019); Schellinger (2021); Trent (2018)

[1] Feedback provided by digital agents will also be considered under the technology-facilitated methods of assessment.

Oral Questioning/Discussion References to teachers' use of questioning/discussion as a tool to determine what learners know. This includes conferencing.	Brookhart (2020); Chin (2007); Cizek (2019); Hattie (2007); Hattie (2019); Lysaght (2019); McMillan (2020); Murchan (2017); NCCA (2007); Roshenshine (2012); William (2015)	Baas (2020); Chen (2022); Chang (2020); Coe (2019); Collins (2018); Education Endowment Foundation (EEF) (2021a); Harris (2022); Heitink (2016); Heritage (2013); Hodgen (2020); Klute (2017); Pyle (2017); Schildkamp (2020); Zhang (2019)	Aguirre-Munoz (2021); Bartels (2019); Brough (2012); Björklund (2017); Calder (2013); Cervetti (2012); Cunnington (2014); García-Carrillo (2021); Hammond (2017); Lehrer (2021); Liston (2018); López-Leivaa (2016); Luna (2015); Maitles (2012); Mildenhall (2021); Miller (2019); Schellinger (2021); Speldewinde (2022); Trent (2018); Tytler (2021)
Performance-Based Assessments[2] Any presentation of work by the learner to demonstrate their learning e.g. creating a model/artefact, concept map, giving a presentation, performances. The collation and presentation of *multiple* work samples to demonstrate their learning i.e. portfolios are also included here.	Briggs (2019); Brookhart (2015); Brookhart (2020); Cizek (2019); Lysaght (2019); McMillan (2020); Murchan (2017); NCCA (2007); William (2015)	Alqassab (2023); Bartholomew (2019); Baas (2020); Chang (2020); Chen (2022); Cutumisu (2019); Graham (2015); Harris (2022); Hodgen (2020); Killian (2021); Klute (2017); Lane (2020); Long (2022); Li (2020); Miller-Bains (2017); Pyle (2017); Tang (2020); Wheadon (2020); Zhang (2019); Zheng (2020)	An (2013); Baptiste (2022); Birsa (2018); Bazemore (2015); Brough (2012); Brugar (2012); Cannon-Ruffo (2020); Cassidy (2022); Cunnington (2014); Edsall-Giglio (2012); Ensign (2017); Fitzpatrick (2018); Fragakis (2019); Gerke (2017); Halimah (2021); Hardiman (2019); Jordan (2016); Jia (2021); Kok (2014); LaMotte (2018); Levy (2018); Liston (2018); Maitles (2012); McDowall (2019); Mildenhall (2021); Moss (2019); Öztürk Yılmaztekin (2017); Quigley (2019); Revelle (2019); Rule (2012); Savage (2016); Speldewinde (2022); Tank (2014); Trent (2018); Trinter (2021); Vacca (2022); Vallera (2016)
Rubrics/Shared Success Criteria The use of a rubric or a set of shared success criteria (by learners, teachers or researchers) to assess the quality of work undertaken e.g. rubric on writing.	Brookhart (2015); Brookhart (2020); Cizek (2019); Lysaght (2019); McMillan (2020); Murchan (2017); NCCA (2007); William (2015)	Bartholomew (2019); Double (2019); Graham (2015); Harris (2022); Heitink (2016); Hodgen (2020); Karaman (2021); Kennedy (2022); Killian (2021); Klute (2017); Miller-Bains (2017); Panadero (2013); Tang (2020); Wheadon (2020); Zheng (2020)	Bazemore (2015); Brough (2012); Brugar (2012); Cunnington (2014); Duke 92021); Fitzpatrick (2018); Fragakis (2019); Frankel (2015); Gerke (2017); Hammond (2017); Jia (2021); Jordan (2016); Kok (2014); Levy (2018); McDowall (2019); Moss (2019); Sáez-López (2016); Trent (2018); Vacca (2022); Vallera (2016)
Self-Assessment Any instances where learners evaluated their work or experiences in a reflective way e.g. self-assigning a mark, self-identifying areas of strength or further development.	Brookhart (2020); Chen (2021); Lysaght (2019); McMillan (2020); Murchan (2017); Wiliam (2015)	Andrade (2019); Double (2019); Education Endowment Foundation (EEF) (2021a); Graham (2015); Hartmeyer (2018); Heitink (2016); Karaman (2021); Kenae (2017); Killian (2021); Klute (2017); Lane (2020); Lee (2020); Sanchez (2017); Schildkamp (2020); Zheng (2020)	Brough (2012); Cunnington (2014); Fragakis (2019); Gerke (2017); Hammond (2017); Jordan (2016); LaMotte (2018); Lovemore (2015); Maitles (2012); Mård (2022); McDowall (2019); Moss (2019); Quigley (2019)
Standardised Assessments	Brookhart (2020); Hoover (2013); Lysaght (2019); McMillan (2020); Murchan (2017); NCCA (2007);	Brookhart (2016); Collins (2018); Cutumisu (2019); Graham (2015); Koenka (2021); Klute (2017); Lane (2020);	Atalay (2015); Bazemore (2015); Cannon-Ruffo (2020); Casady (2015); Doyle (2014); Cunnington

[2] It should be noted that assessments exist along a continuum based on how much construction is required by learners (Brookhart, 2015). This makes it difficult to clearly delineate between different assessment practices. For example, essays and written products are often considered performance based assessments. However, they can also be used in teacher-designed or state tests. For the purposes of this report, and in line with Brookhart (2015), when an assessment such as an essay is "administered in a context that is not a test, it is treated as a performance assessment" (p. 3/4). If an essay is to be done during a designated testing time, they can be considered test questions.

Any assessments involving formal, standardised administration or scoring procedures e.g. state tests of literacy or numeracy, cognitive assessments, psychological assessments, benchmark tests	O'Leary (2019)	Long (2022); Pyle (2017); Sanchez (2019)	(2014); Fazio (2019); Feldwisch (2014); Frankel (2015); Graham (2016); Gray (2022); Harris (2019); Inoa (2014); Lamb (2015); Talbert (2019); McDowall (2019); O'Neal (2017); Panagopulos (2015); Peppler (2014); Muchan (2017); Schugar (2017); Smith-Gayle (2014); Snyder (2014); Tucker (2017); White (2014); Wright (2017)
Technology-Facilitated Assessment Use of digital technology to support the assessment process e.g. provision of peer feedback, creation of artefacts using online tools, digital game-based assessment	Brookhart (2020); Chen (2021); Cizek (2019); Davidson (2010); Hattie (2007); Hattie (2019); Lysaght (2019); McMillan (2020); Murchan (2017); Russell (2020)	Alqassab (2023); Bartholomew (2019); Bolat (2022); Caballero-Hernández (2017); Chen (2022); Chang (2020); Cutumisu (2019); Double (2019); Graham (2015); Hartmeyer (2018); Harris (2022); Heitink (2016); Hodgen (2018); Killian (2021); Klutes (2017); Lee (2020); Li (2020); Merritt (2022); Murchan (2017); Pyle (2017); See (2021); Tang (2020); Zhang (2019)	Cannon-Ruffo (2020); Ensign (2017); García-Carrillo (2021); Mård (2022); Monteiro (2021); Sáez-López (2016)

ERIC (via EBSCOHost)
Date: 06/03/2023
Searcher: Paula Lehane

#	Searches	Results
S1	**TI** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	53,566
S2	**AB** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	223,429
S3	**KW** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	0
S4	**S1** OR **S2** OR **S3**	235,428
S5	**S4** AND "primary school" OR "elementary school" OR "primary education" OR "elementary education" OR "middle school" OR "middle edu*" OR "early child* edu" OR "early years edu*"	69,097
S6	**S5** AND "systematic review" OR "meta-analysis" OR "best evidence"	501
S7	**Limiters applied to S6:** English language; 01-01-2012 - present; peer-reviewed	278

Articles brought forward for inclusion from this search (based on Title/Abstract Screening by Paula Lehane): *n*=26

Education Research Complete (via EBSCOHost)
Date: 06/03/2023
Searcher: Paula Lehane

#	Searches	Results
S1	**TI** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	57,615
S2	**AB** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	264,929
S3	**KW** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	28,734
S4	**S1** OR **S2** OR **S3**	285,638
S5	**S4** AND "primary school" OR "elementary school" OR "primary education" OR "elementary education" OR "middle school" OR "middle edu*" OR "early child* edu" OR "early years edu*"	18,000
S6	**S5** AND "systematic review" OR "meta-analysis" OR "best evidence"	217
S7	**Limiters applied to S6:** English language; 01-01-2012 - present; peer-reviewed	177

Articles brought forward for inclusion from this search (based on Title/Abstract Screening by Paula Lehane): *n*=13

Web of Science (Clarivate)
Date: 06/03/2023
Searcher: Paula Lehane

#	Searches	Results
S1	**TI** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	948,673
S2	**AB** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	4,476,980
S3	**KW** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	271,448
S4	**S1** OR **S2** OR **S3**	5,047,120
S5	**S4** AND "primary school" OR "elementary school" OR "primary education" OR "elementary education" OR "middle school" OR "middle edu*" OR "early child* edu" OR "early years edu*"	265,025
S6	**S5** AND "systematic review" OR "meta-analysis" OR "best evidence"	27,391
S7	**Limiters applied to S6:** English language; 01-01-2012 - present; peer-reviewed; indexed under 'Educational/Education Research'	420

Articles brought forward for inclusion from this search (based on Title/Abstract Screening by Paula Lehane): *n*=19

Scopus (via Elsevier)
Date: 06/03/2023
Searcher: Paula Lehane

#	Searches	Results
S1	**TITLE-ABS-KEY** "assess*" OR "classroom assess*" OR "formative assessment" OR "summative assessment" OR "assessment for learning" OR "assessment of learning"	201,819
S2	**S1 AND** "primary school" OR "elementary school" OR "primary education" OR "elementary education" OR "middle school" OR "middle edu*" OR "early child* edu" OR "early years edu*"	5,596
S3	**S2 AND** "systematic review" OR "meta-analysis" OR "best evidence"	1,660
S4	**Limiters applied to S6:** English language; 01-01-2012 - present; peer-reviewed; indexed under 'Social Sciences' or 'Arts and Humanities' or 'Multidisciplinary'	501

Articles brought forward for inclusion from this search (based on Title/Abstract Screening by Paula Lehane): *n*=14