A Systematic Literature Review on the Effectiveness of Universal Design for Learning in Second-level Education

Ann Devitt[1], Joanne Banks[1], Aibhin Bray[1], Sergio Sanchez Fuentes[2], Marta Sandoval, Katerina Riviou[3], Margaret Flood[4], Jean Reale[5], Darren Byrne[6], Eimear McCarthy, Silvia Terrenzio[6]

[1] *School of Education, Trinity College Dublin, the University of Dublin, Ireland, Ann.Devitt@tcd.ie*
[2] Faculty of Teachers Training and Education, Autonomous University of Madrid, Spain
[3]Research & Development Dept., Ellinogermaniki Agogi, Athens, Greece
[4]Maynooth University, Co. Kildare, Ireland
[5]Mary Immaculate College, Limerick, Ireland
[6]St Joseph's Secondary School, Rush, Dublin, Ireland
[6]CVO Vitant, Antwerp, Belgium

ABSTRACT

Education systems around the world are grappling with increasingly diverse student populations and a focus on the right to education for every learner. New innovative pedagogies such as Universal Design for Learning (UDL) have emerged as a response to more traditional approaches to teaching and learning. While much of the research on UDL to date has been in further and higher education, less is known about its effectiveness in second-level education.

This Systematic Literature Review provides an analysis of the effectiveness of UDL in 22 empirical studies from 1200 papers screened and interrogates the aspects of the UDL framework commonly deployed in interventions. The review indicates growth in the evidence base for UDL effectiveness in second-level education. The findings confirm positive effects for UDL interventions on learners' academic and social outcomes. The paper highlights that the aspects of UDL within the teachers' control are well addressed (Representation), whereas elements that aim to build learner autonomy and self-regulation are less evident in the interventions.

The results of this literature review provide the first systematic insight into the effectiveness of UDL in mainstream second-level education. While many positive outcomes were identified, this paper highlights the need to acknowledge existing pedagogical approaches that could support the design of UDL implementations. In particular, teaching and learning strategies such as formative assessment, culturally responsive teaching and multilingual education, and autonomy-supportive pedagogy, could usefully be integrated within a broader UDL framework.

Keywords

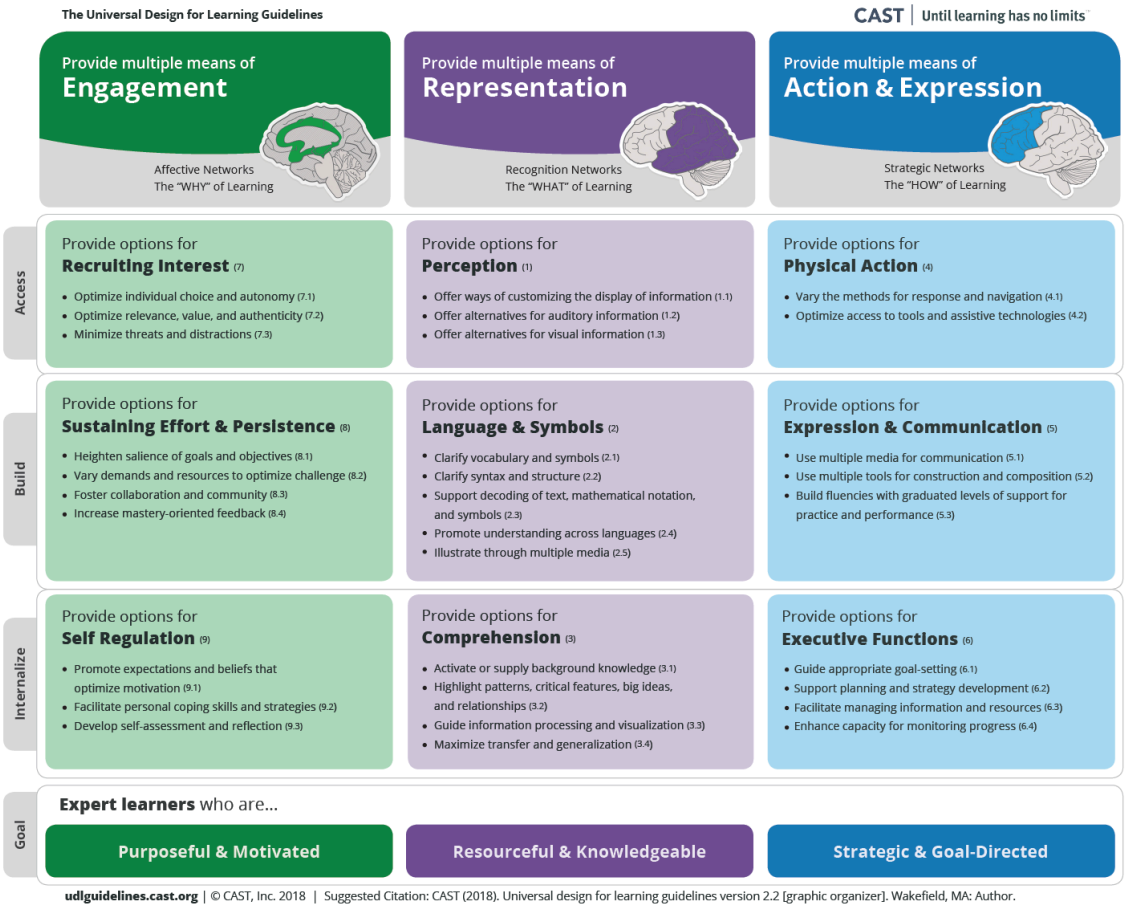
Universal Design for Learning, inclusive education, pedagogy, systematic literature review, second-level education, K-12 education

INTRODUCTION

Inclusive education has become a key priority of education and human rights policy. International organisations such as UNESCO have made it a policy priority with the Education 2030 Framework for Action (UNESCO, 2016), which seeks to provide a framework for countries wishing to implement Sustainable Development Goal 4, prioritising inclusive and equitable education for all (UNESCO, 2016). Across different national contexts, inclusive education is a significant policy issue, with many countries attempting to respond to increasingly diverse student populations with existing, historical, systems of 'special education' (Boyle & Anderson, 2020; Grynova & Kalinichenko, 2018; Schuelka, 2018). Over the last two decades, however, some countries have begun moving towards innovative pedagogies such as Universal Design for Learning (UDL), with the aim of responding to learner diversity and creating more inclusive school environments for everyone (Galkienė & Monkevičienė, 2021). Developed in the 1980s by the U.S. non-profit organisation CAST, the UDL framework seeks to move away from a 'one size fits all' curriculum approach to one where learners are provided with choice and flexibility about how they learn and how they demonstrate what they have learned (Chardin & Novak, 2021). UDL facilitates inclusive learning environments by assuming student variability and removing unnecessary barriers in the learning process from the outset.

The UDL framework is comprised of three principles (Multiple Means of Engagement, Action & Expression and Representation), nine guidelines and 30 checkpoints (detailed in Figure 1), and has its foundations in empirical research from disciplines including neuroscience, learning sciences and cognitive psychology (CAST, 2022). There is a significant body of work and discussion on UDL's capacity to respond to student diversity and create more equitable access to the curriculum for every learner (Chardin & Novak, 2021). Studies show: the potential impact of UDL in creating more inclusive settings in higher education (Fovet, 2020, 2021a); the barriers to UDL implementation in higher education (Fovet, 2021b); teacher perspectives on UDL implementation (Capp, 2020); the application of UDL for students with 'special educational needs' (Scott & Temple, 2017) or disabilities (Lowrey & Smith, 2018; Seok, DaCosta, & Hodges, 2018); and much broader literature on the potential of UDL for diverse learners (Chardin & Novak, 2021; Fovet, 2019).

***Figure 1**: Universal Design for Learning Guidelines (udlguidelines.cast.org)*



Despite this strong evidence base for the development of the framework within neuroscience, there is a lack of research highlighting the effectiveness of UDL for learners across the different education sectors (Al-Azawei, Serenelli, & Lundqvist, 2016; Ok, Rao, Bryant, & McDougall, 2017), with some researchers raising concerns that UDL is being introduced to education policy without sufficient evidence on its effectiveness (Murphy, 2021). This gap in research is particularly evident in second-level education. Given the growing emphasis on UDL in the context of inclusive education worldwide (Katz, 2013; Loreman, 2017), the aim of this paper is to examine the evidence base for UDL within the context of second-level education.

The effectiveness of UDL: what do we know?

It is important to note that several key Systematic Literature Reviews (SLRs) have already been undertaken on the effectiveness of UDL on learner outcomes across the different sectors of education. AlRawi and AlKahtani (2021), Seok et al. (2018), Capp (2017) and Mangiatordi and Serenelli (2013) have all examined the empirical evidence

of the effectiveness of UDL across educational settings. Overall, the findings of these studies suggest positive effects but emphasise the lack of empirical evidence on the effectiveness of UDL and highlight that much of the available research comes from North America. Using a meta-analysis approach of 80 articles on the evidence for the effectiveness of UDL, Mangiatordi and Serenelli (2013) found that only a minority of abstracts examined showed a positive impact of UDL, but they also note the growing information on the topic at that time. They also highlight the extent to which the UDL literature focuses on position and opinion papers and how they are predominantly focused on the United States (Mangiatordi & Serenelli, 2013). In 2017, Capp (2017) published the findings of a meta-analysis of 18 articles that examined the effectiveness of UDL using pre- and post-testing scores of students. He found that UDL is an effective teaching methodology for improving the 'learning process' for all students.

Also in 2017, Ok et al. (2017) examined the extent to which UDL instruction impacted the academic and social outcomes of students in pre-K to grade 12. They found that UDL increased engagement and access to general education for students with disabilities and improved the academic and social outcomes of all students. They do however note the efficacy of UDL studies varies considerably across the different interventions. Focusing specifically on students with intellectual disabilities, AlRawi and AlKahtani (2021) explore the educational impact of UDL, finding seven articles that indicate a positive impact on the academic, social, and behavioural outcomes of the students involved. Also using a disability focus, Seok et al. (2018) examined the degree to which UDL methods are effective for students with and without disabilities. Their findings show that 15 of the 17 studies included in the review reported 'effective outcomes'.

Methodology

This paper builds on the meta-analyses and SLRs discussed above by updating the findings with the most recent evidence-based research and providing a finer-grained analysis of UDL implementations. In line with the prior work, this study identifies gaps in the extant literature with a view to influencing research going forward. This study takes a Systematic Literature Review as our methodological approach as it systematically identifies, analyses and critically synthesises existing publications on a chosen topic to provide an insight into the 'body of evidence' available in order to influence policy and practice (Newman & Gough, 2020; Zawacki-Richter, Kerres, Bedenlier, Bond, & Buntins, 2019). The study adopted a Cochrane systematic review approach (Higgins & Green, 2011), using the EPPI-Reviewer tool (eppi.ioe.ac.uk) and taking the following steps: 1) specify review research questions; 2) define inclusion/exclusion criteria; 3) design and apply search strategy; 4) screen articles for inclusion; 5) extract study data and conduct quality appraisal; 6) complete qualitative synthesis to address the study research questions.

Research Questions

This systematic review is situated within a broader Erasmus+ research study on UDL in second-level education, involving a transnational European consortium. For this reason, the primary focus for the systematic review is on education at second level including young people from approximately age 11 to the end of second-level schooling (age 18 or 19). The review seeks to explore the evidence base for UDL in all school settings at this age range and to identify aspects of UDL that are well established, as well as gaps in implementation and research within the field (Polanin, Maynard, & Dell, 2017). The systematic review therefore will address the following research questions:

(1) What is the evidence that UDL implementations positively influence student outcomes in second level education?
(2) What aspects of UDL are implemented in studies at second level?

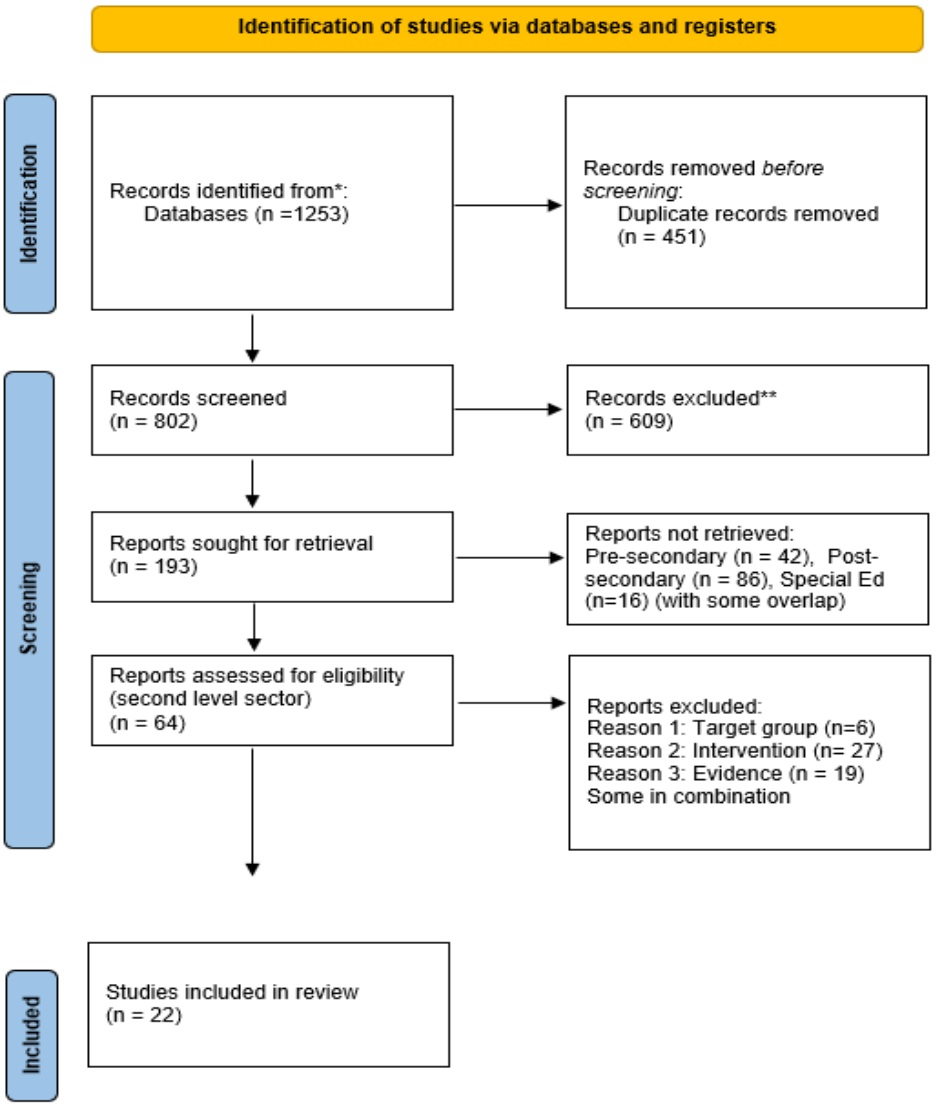
Inclusion Criteria

Prior to beginning article search and screening, the Inclusion/Exclusion criteria **(Table 1)** were agreed within the research team. Given the relatively recent emergence of UDL in education, no time limit on publications was included. Inclusion criteria 4 and 5 were only implemented at full text screening. The details of the search and screening process are set out in the PRISMA diagram **(Figure 2)**.

TABLE 1: *Inclusion/exclusion criteria*

Inclusion	Exclusion
1. Published in English	Published in languages other than English
2. Peer-reviewed publications	Grey literature or non-peer-reviewed publications
3. Includes an empirical study of UDL	Descriptive, theoretical or review papers that do not include an empirical study
4. Focuses on learners in second level education (approximately aged 11-19)	Focuses on pre-K, elementary/primary learners, or learners in further or higher education
5. Includes a focus on student outcomes (including academic, social, affective, or other outcomes)	No focus on student outcomes of any kind

Figure 2: *PRISMA Flow Chart of Systematic Review process*



Search Strategy

The initial search strategy was designed to be as open and inclusive as possible. The search string therefore included only the terms "Universal Design for Learning" or "UDL" in the title or abstract, with the search conducted across five databases: ERIC, Applied Social Sciences Index & Abstracts (ASSIA) ProQuest, JSTOR and PsycInfo. The results of the databases searches were validated with reference to previous UDL review papers. The refined results of the search led to identification of 823 articles, which were imported into EndNote and EPPI Reviewer for screening.

Screening

The article screening process proceeded in two phases with the author team. Initial title and abstract screening was conducted by five of the author team beginning with a pair-wise cross-validation process of 10% of the articles to ensure consistency. During this phase, all included articles were also coded for education level (pre-K, primary, secondary, further/higher education). In total 193 articles met the inclusion criteria, of which 64 focused on secondary education. Eight members of the author team conducted the full-text screening of the sixty-four second-level articles resulting in 22 articles for inclusion in the review.

Data Extraction, Appraisal and Analysis

In line with UDL reporting guidelines (Rao et al., 2018), all key methodological, participant and outcome data were extracted from the review articles. As regards the UDL principles and checkpoints enacted in the study interventions, a small number of papers reported this in a very precise and fine-grained manner. However, for the most part, this was deduced from a focused content analysis of each study's description of the intervention materials and process. Quality Appraisal tools available within EPPI Reviewer were used to explore the quality of evidence in the review articles. However, given the small number of articles identified in the screening process, no articles were excluded on the basis of this appraisal. Taking the Weight of Evidence (Gough, 2007) approach, the final synthesis addresses the extent to which the review articles contribute evidence to address the study research questions.

Findings

This section provides an overview of the characteristics of the 22 papers followed by a synthesis of the findings under the study research questions.

Research methods

The reviewed articles used a range of methodological approaches to examine the effectiveness of UDL at second level education with quantitative (10), qualitative (8) and

mixed methods (7) all identified. In relation to quantitative approaches, mainly non-randomised intervention studies are used (Daley, Xu, Proctor, Rappolt-Schlichtmann, & Goldowsky, 2020; King-Sears & Johnson, 2020; King-Sears et al., 2015; Scott et al., 2011). There are also two randomized pre/post-test designs (Kennedy, Thomas, Meyer, Alves, & Lloyd, 2014; McMahon, Wright, Cihak, Moore, & Lamb, 2016), the latter using generalised Latin squares. Two pre/post-test intervention designs without control groups were also noted (Marino, 2009; Marino, Black, Hayes, & Beecher, 2010). A number of mixed methods studies provided a rich mix of a quasi-experimental study with a qualitative perspective drawn from interviews and focus groups with teachers or students as well as other artefacts or observations of learning (Hitchcock, Rao, Chang, & Yuen, 2016; Katz, 2013, 2015; Marino et al., 2014; Smith, Lowrey, Rowland, & Frey, 2020). Ender et al.'s (2007) mixed methods study used a variety of data collected at project, school and classroom level including a range of outcome reports and artefacts. The most commonly used qualitative methodological approach was case study (Dymond et al., 2006; Robinson, 2017; Staats & Laster, 2018). Basham, Meyer, and Perry (2010) took a Design-Based Research approach in which qualitative information was collected through observations, video recordings, diaries and interviews and Katz and Sokal (2016) report on a qualitative aspect of their broader mixed-methods study (Katz, 2013, 2015).

Scale and Duration

There was substantial diversity in scale and scope between the various studies, with numbers of participants ranging from n=5 to n=1153 (Table 2). Unsurprisingly, studies that involved more than one school/setting in the study tended to have higher numbers of participants. Similarly, the duration of the interventions varied widely, ranging from one 90-minute lesson to one year or more. The majority of studies however (13/22), are of significant scale in both number of participants and duration (Table 2), which strongly suggests that the evidence base for effective use of UDL at second level is substantial.

***TABLE 2**: SLR Study Scale and Duration*

	DURATION				
SCALE	Total	Very short (1 lesson)	Short (under 4 weeks)	Medium (4 weeks/1 unit or over)	Long (12 weeks or over)
	22	3	4	7	8
Very small (n<15)	2		1		1
Small (16<n<35)	2		2		

Medium (35<n<100)	3	1	1		1
Large (101<n<500)	8	1		2	5
Very large (n>500)	7	1		5	1

Study Settings, Participants and Curriculum Areas

All articles included in this review were from North America, with 19 from the United States and 3 from Canada. Thirteen of the studies involved more than one school. Two studies focused specifically on special education or resource classrooms and another two were in out-of-school settings. In many of the studies, the learners were classified according to gender, ethnicity, socio-economic status (using free/reduced price lunch as a proxy), and/or learning (dis)ability. Fifteen of the studies also noted teacher participants. Diverse curriculum areas were addressed in the studies, with nine mentioning STEM subjects (Mathematics, Astrophysics, Chemistry, and Biology). Four studies addressed History and Social Sciences and three explore literacy skills such as reading and writing. In the rest of the studies no specific curriculum area was mentioned. All of the studies are summarised in Appendix 1.

Evidence Base for UDL

RQ1: *What is the Weight of Evidence (Gough, 2007) that UDL implementations positively influence student outcomes in second level education?*

Although the 22 articles share a common objective of examining a UDL intervention in second-level education, there was much variation in the intended outcomes of the research. The majority of the studies (n=13) explore students' academic outcomes, four in the area of language and literacy (Daley et al., 2020; Hitchcock et al., 2016; Kennedy et al., 2014; Smith et al., 2020) and five in the area of STEM (King-Sears & Johnson, 2020; King-Sears et al., 2015; Marino, 2009; Marino et al., 2010; Marino et al., 2014). Three articles report on students' social or affective outcomes (Abell, Jung, & Taylor, 2011; Dymond et al., 2006; Kortering, McClannon, & Braziel, 2008). One is unspecific in what kinds of outcomes are addressed (Robinson, 2017). The remaining five address both academic and social or affective outcomes. These include the three studies on the Three Block Model as well as the studies by Scott et al. (2011) as well as Staats and Laster (2018). The studies that include a pre/post-test design use specific outcome measure to determine student academic or other outcomes, in some cases with an additional qualitative perspective from other data sources. Several papers report student outcomes

as reported by either the teacher or the student, rather than based on a specific outcome measure **(Table 3)**.

***TABLE 3**: Summary of articles in SLR*

Author	Scale	Disability focus	Method	Evidence	Student outcome
Katz (2013)	V large	Included	Mixed	Positive	Academic and social outcomes
Katz (2015)	V large	No	Mixed	Positive	Academic and social outcomes
Katz and Sokal (2016)	Large	Included	Qual	Positive	Academic and social outcomes
Scott et al (2011)	V small	Yes	Quant	Positive	Academic and social outcomes
Staats and Laster (2018)	V large	No	Qual	Positive	Academic and social outcomes
Basham et al (2010)	Small	Included	Qual	Positive	Academic outcome
Daley et al (2016)	Large	Included	Quant	Positive	Academic outcome
McMahon et al (2016)	Medium	Included	Quant	Positive	Academic outcomes
Daley et al (2020)	Large	Yes	Quant	None	Academic outcomes
Hitchcock et al (2016)	Medium	Included	Mixed	Positive	Academic outcomes
Kennedy et al (2014)	Large	Yes	Quant	Positive	Academic outcomes
Smith et al (2020)	V large	Yes	Mixed	Positive	Academic outcomes
Ender et al (2007)	Large	Included	Mixed	Positive	Academic outcomes
King-Sears et al (2015)	Medium	Included	Quant	Mixed	Academic outcomes
King-Sears and Johnson (2020)	Small	Included	Quant	Positive	Academic outcomes
Marino (2009)	V large	Yes	Quant	Mixed	Academic outcomes
Marino et al (2010)	V large	Yes	Quant	Mixed	Academic outcomes
Marino et al (2014)	Large	Yes	Mixed	Mixed	Academic outcomes
Smith et al (2014)	NA	NA	Qual	None	Unspecified
Abell et al (2011)	V large	No	Quant	Positive	Social outcome

Dymond et al (2006)	Large	Included	Qual	Positive	Social outcome
Kortering et al (2008)	Large	Yes	Quant	Positive	Social outcome
Robinson (2017)	V small	Yes	Qual	Positive	Unspecified

Given the variety within the literature, this analysis focused broadly on the extent to which the papers provide an evidence base that UDL implementations are effective at second level, regardless of study objectives or outcomes under examination. The findings show that of the 22 articles selected for inclusion in the SLR, 16 found positive, or somewhat positive, effects of UDL on learners' academic or social outcomes, four articles reported mixed findings, and two showed neutral or no effects. While learners with disabilities are the focus of some studies, the majority were full class interventions, and the results demonstrate the potential usefulness of UDL for a wide range of students.

The articles included in the SLR include several large studies that provide solid evidence for the effectiveness of UDL on student social and/or academic outcomes. Of particular note, are a series of substantive studies by Katz (2013, 2015), Katz and Sokal (2016) as well as King-Sears & Johnson (2020), King-Sears et al. (2015), Marino, (2009), Marino et al. (2010), Marino et al. (2014) and Hitchcock (2016), which examined various aspects of student learning with intervention and control groups, mixed methodologies and pre/post-test analysis. These studies report results that highlight positive findings overall on outcomes that include test scores, reading comprehension, social interactions, and self-direction. These positive results appear consistent when comparing the effectiveness of the UDL interventions for students with and without disabilities.

Not surprisingly, some of the literature includes a focus on the use of digital technologies with, or as a part of, the UDL interventions (see Bray et al., 2024). Fourteen of the studies included in this review made specific reference to technology, with eight exploring bespoke technologies designed according to UDL principles, three making use of existing digital tools to facilitate a UDL approach, and three broadly referencing how technology can be used to support UDL.

In summary, the systematic review found that there is solid evidence for effectiveness of UDL implementation on student outcomes, including academic measures of reading, writing, STEM, and social outcomes. Across the range of outcomes explored, positive outcomes have been demonstrated for students with disabilities as well as for the student body as a whole. Given that Engagement is one of the three core principles of UDL, it is surprising that student engagement as an outcome is only dealt with in one

study (Kortering et al., 2008). However, although the articles provide evidence across a broad range of curriculum areas and student outcomes, all the studies in this review took place in North America; the extrapolation of this evidence base to other cultural contexts will require careful consideration. In addition, while some of the studies are large-scale, none of them are at a national level or provide longitudinal data.

UDL Aspects Implemented

RQ 3: *What aspects of UDL are implemented in studies at second level?*

Given the scope of UDL across teaching and learning, it is difficult to address all aspects of the framework within a single empirical study. This SLR set out to determine the distribution of empirical research across UDL principles, guidelines, and checkpoints (see **Figure 1**). **Table 4** maps the review studies to the UDL principles implemented. Three studies will not be discussed here as they are presented as generic UDL implementations without referencing specific UDL principles (Abell, Jung, & Taylor, 2011; Ender et al., 2007; Kortering et al., 2008). Five articles cover comprehensive implementations of UDL, including three articles exploring two distinct studies of the 3-Block Model (Katz, 2013, 2015; Katz & Sokal, 2016), the TeenACE intervention (Hitchcock et al., 2016) and Robinson's (2017) library case-study.

TABLE 4*: UDL Principles and Guidelines in Focus in Articles*

	Engagement (n=13)	**Representation (n=14)**	**Action and Expression (n=11)**
Access (14)	Recruiting Interest (n=13) Daley et al. (2020) Dymond et al. (2006) Kennedy et al. (2014) King-Sears et al. (2015) King-Sears and Johnson (2020) Marino (2009) Marino et al. (2010) Marino et al. (2014) Staats and Laster (2018)	Perception (n=14) Basham et al. (2010) Dymond et al. (2006) McMahon et al. (2016) Scott et al. (2011) Staats and Laster (2018)	Physical Action (n=2) McMahon et al. (2016)

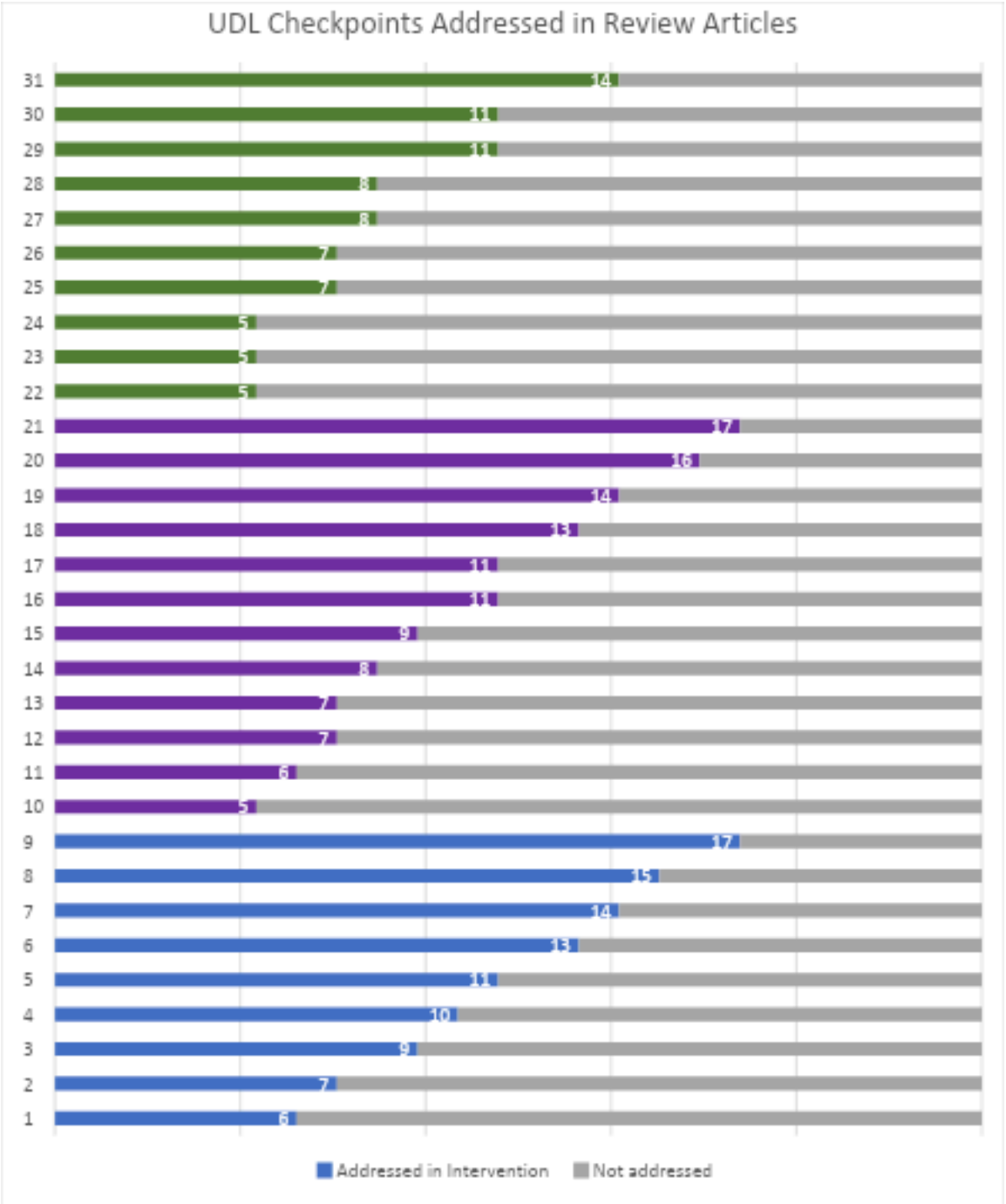
Build (11)	Sustaining Effort and Persistence (n=8) Daley et al. (2020) Marino et al. (2010) Marino et al. (2014) Staats and Laster (2018)	Language and Symbols (n=9)	Expression and Communication (n=11) Basham et al. (2010) Daley, Hillaire, and Sutherland (2016) Daley et al. (2020) King-Sears et al. (2015) King-Sears and Johnson (2020) Marino (2009) Marino et al. (2010) Marino et al. (2014) Smith et al. (2020) Staats and Laster (2018)
Internalise (12)	Self-Regulation (n=4)	Comprehension (12) Basham et al. (2010) Dymond et al. (2006) Staats and Laster (2018)	Executive Functions (n=11) Basham et al. (2010) Daley et al. (2016) Daley et al. (2020) King-Sears et al. (2015) King-Sears and Johnson (2020) Marino (2009) Marino et al. (2010) Marino et al. (2014) Smith et al. (2020) Staats and Laster (2018)
Access, Build & Internalise	Engagement overall (4) Basham et al. (2010) Daley et al. (2016) Dymond et al. (2006) Scott et al. (2011)	Representation Overall (9) Daley et al. (2016) Daley et al. (2020) Kennedy et al. (2014) King-Sears et al. (2015) King-Sears and Johnson (2020) Marino (2009)	Action and Expression overall (1) Scott et al. (2011) - in Intervention II

		Marino et al. (2010) Marino et al. (2014) Smith et al. (2020)	
ALL (5)	Hitchcock et al. (2016) (Katz, 2013, 2015); Katz and Sokal (2016): the 3-Block Model Robinson (2017)		
Not specified (3)	Abell et al. (2011): classrooms consistent with UDL, not intentional UDL design Ender et al. (2007): school level change across State not specified Kortering et al. (2008)		

Of the remaining fourteen studies, which will be the focus of this section, most studies address explicitly or implicitly between two and seven UDL principles. All fourteen articles address Multiple Means of Representation principle at a minimum. Specifically, the Perception guideline is addressed to some degree in all of the articles. The Comprehension guideline is also well addressed (n=12 articles). As regards Multiple Means of Engagement, almost all (n=13) studies address this domain explicitly, all of which focus (at least) on Recruiting Interest, sometimes relying on the novelty effect of technology to achieve this. Fewer studies (n=8) maintain an explicit focus on Sustaining Effort & Persistence, often relying on collaborative activity (checkpoint 8.3) to achieve this goal. Only four articles address Self-Regulation, and all of these deploy explicit scaffolded strategies to address Executive Function and Self-Regulation simultaneously. All of the eleven papers with a focus on Action & Expression explicitly addressed the guidelines of Expression & Communication and Executive Functions often through scaffolded expression and strategy support. Only two articles focus on providing options for Physical Action, with McMahon et al. (2016) focusing exclusively on this guideline in conjunction with offering multiple options for Perception.

An analysis of the UDL checkpoints addressed in the interventions provides a more fine-grained account of the aspects of UDL that are most frequently deployed in UDL implementations at second level, see **Table 5**.

TABLE 5: *Summary of UDL Checkpoints in Focus in Articles*



This highlights that UDL implementations focus predominantly on offering choice (checkpoint 7.1), particularly through alternative audio and visual representations of content (checkpoints 1.2 and 1.3) as well as introducing strategies of different kinds to support learning (checkpoint 6.2). The organisation of content in a structured way to support comprehension is also relatively commonly addressed, with graphic organisers

as the core tool to highlight patterns (checkpoint 3.2) and guide information processing (checkpoint 3.3). Integrating new learning is less frequently addressed whether through accessing prior knowledge (checkpoint 3.1) or by relating to other contexts (checkpoint 3.4). In some papers, Problem-Based Learning is deployed to achieve this as well as heightening the relevance of content for learners. Within the principle of Representation, there is a clear gap in the literature in relation to Language & Symbols, in particular when it comes to supporting multilingual learners. As regards Action & Expression, there are few studies that explicitly include options for Physical Action. This may be due to the characteristics of the student cohorts that may or may not need alternatives here, or it could relate to the availability of resources within the school settings. Finally, there is a clear gap across the UDL checkpoints that relate to supporting affective and cognitive dimensions of self-directed learners (checkpoints 8.4, 9.1, 9.2, 9.3, 6.1, 6.4). The overall profile of UDL implementations is that for the most part those aspects that are within teachers' control are quite well addressed while those that relate to building student capacity to integrate and direct their own learning are less so.

Discussion

This systematic review has expanded on previous reviews, to provide an up-to-date and fine-grained analysis of empirical research on UDL at second level. The findings of the SLR indicate the emergence of a strong evidence base for the effectiveness of UDL, in North America at least, at this level of education, with a number of large-scale and rigorous studies using a range of methodological approaches. In particular, the large-scale, mixed-methods studies provide an in-depth view of student outcomes. The majority of the studies focus on academic outcomes alone, including standardised test scores. Measures of student engagement, well-being, self-efficacy, or self-concept measures have not been widely used in the field to date. The holistic nature of UDL however, lends itself to holistic outcome measures. Furthermore, the voice of the learners is quite limited in existing research and would greatly enrich the evidence base in this very learner-focused field.

As it stands, different studies are measuring different outcomes, and there is no broad-based indicator of outcomes that allows for comparison across studies, comparable to Booth and Ainscow's (2002) Index for Inclusion. Such an indicator might also be useful in considering fidelity of UDL implementations, which was dealt with in some papers using a range of instruments. An instrument comparable to the UDL Scan Tool (Smith & Harvey, 2014), which is used to understand how well technology aligns with the framework, could be helpful to examine UDL in educational settings.

The fine-grained analysis of existing studies at the level of individual UDL checkpoints, identifies a leaning towards options for Representation and, to a somewhat lesser extent, options for Action & Expression. It highlights strategies or pedagogical approaches that have been used successfully within the evidence base. It also identifies

significant gaps in UDL implementations. We have found that those aspects that are within the control of the teacher have been well addressed in the articles included. However, when it comes to building capacity in the learner, these aspects are less thoroughly implemented. This discussion highlights some particular areas that require consideration within the field.

The identified gap in implementation in relation to the Language & Symbols guideline, particularly the limited provision of support for multilingual learners, is highly problematic. Literacy has a central position within education systems globally, as evidenced by league tables based on national and transnational standardised literacy testing (PISA, PIACC, etc.). In particular literacy plays a key role in high-stakes assessment (Lobascher, 2011). Therefore, it is crucial that UDL implementations offer supports for accessing and producing written texts, as evidenced in the four articles for which literacy outcomes are the measure of success. The implementations certainly provide alternatives to traditional text-based media (e.g., through the use of audio or visuals) or conversion options for text into other media (e.g., Text-to-Speech software) but they are less clear on the supports deployed to access and produce written text directly, especially for multilingual learners. Language technologies have evolved to the point where high-quality, automatic translation, transcription and other tools are readily available, offering learners mechanisms to engage with texts in a range of languages. Indeed, it is possible that the papers do not mention small but crucial tools that may be in use in implementations, such as glossaries, dictionaries, automatic translation, or other easily accessible tools.

Furthermore, there is little discussion of harnessing, valuing or even just acknowledging learners' multilingual repertoires in accessing Multiple Means of Representation or Action & Expression, beyond the language of instruction (e.g., producing dual or home language materials, accessing texts or other materials in learners' home or heritage languages). Given that multilinguals are in the majority globally (Verplaetse & Schmitt, 2010), UDL implementations could usefully engage with the concept and practice of Translanguaging (García & Li, 2014) in order to address this gap.

As noted above, the UDL checkpoints at the intersection of the Self-Regulation and Executive Functions principles are comparatively under-exploited or under-reported in this evidence base. This area links to the domain of Formative Assessment, incorporating the affective and cognitive dimensions of 1) Assessment For Learning, where teachers gather evidence of student learning to direct teaching and learning and to provide feedback, and 2) Assessment As Learning where learners use evidence of progress and feedback to identify the goals and next steps for their own learning (T&L National Forum, 2017). This domain encompasses checkpoints relating to feedback (checkpoint 8.4), self-assessment and reflection (checkpoints 9.3, 6.4) and setting and using learning goals (checkpoints 6.1, 8.1, 9.1). The papers that did tackle these areas

used rubrics or checklists (Dymond et al., 2006), self-check opportunities (Daley et al., 2020; Marino et al., 2014), Self-Regulation Strategy Development (Hitchcock et al., 2016), or self-determined learning model of instruction (Scott et al., 2011). There is extensive literature and clear pedagogical approaches in Assessment For/As Learning and student self-regulation at second level that could be routinely deployed to address these UDL aspects.

Finally, given that UDL often involves the introduction of alternatives, choice and multiple media into a learning context, few studies explicitly address checkpoint 7.3: minimise threats and distractions. Only Kennedy et al. (2014), who proposed a Multimedia Design Framework to address cognitive load in the design of multimedia resources, and Robinson (2017) explicitly tackle this aspect. The other studies do not clearly address how they scaffold and support students to manage the potential distraction of multiple choices, nor do they address whether the kinds of choices provided actually enhance learner autonomy. Indeed, in the literature on autonomy supportive pedagogy (Reeve et al., 2003) it is the provision of action choices rather than option choices that enhance perceptions of autonomy and intrinsic motivation. Within the UDL literature, further critique is required to explore the impact of multiplicity of options, but also the optimum type and scope of options available to learners. As Reeve et al. (2003, p. 389) state, "in practice, the provision for choice is best considered as one contributing element within a larger autonomy-supportive … classroom climate". In particular if not enough emphasis is laid upon articulating and developing learning goals (checkpoints 6.1, 8.1, 9.1) and providing a rationale for choices available, then the provision of choice may not in fact provide an autonomy-supportive environment (Su & Reeve, 2011).

Conclusions

In contrast with previous reviews of UDL (Rao et al., 2014; AlAzawei et al., 2016), this paper notes a growing body of robust evidence for UDL within second level, but they could benefit from greater geographic diversity, more holistic outcome measures and greater scope in terms of time and scale. Perhaps most importantly, there is a need for measurement tools for UDL that are true to the holistic nature of the field but can allow for meaningful comparison across and between studies.

This review also highlights the areas of practice within empirical research, that require further development. Three key areas were identified, including:

(1) Supporting access and production of traditional written text-based media, especially for multilingual learners;
(2) Developing learner autonomy and self-regulation;

(3) Managing choice.

Although these have been identified as gaps through our analysis, this paper argues that the field might benefit from aligning well-understood pedagogical approaches to particular aspects of the UDL framework. A small number of studies in the review corpus demonstrated how it is possible to adapt Evidence-Based Methods (EBMs) from other fields within a UDL framework. To address the identified gaps in UDL implementations, EBMs from formative assessment, culturally-responsive teaching and multilingual education, and autonomy-supportive pedagogy, could usefully be integrated within a broader UDL framework.

Acknowledgements

The UDL-BOE project is co-funded by the Erasmus+ programme of the European Union. This project has been funded with support from the European Commission. Grant number: 2020-1-IE01-KA226-SCH-082934

Declaration of Interest

The authors report there are no competing interests to declare

References

Abell, M. M., Jung, E., & Taylor, M. (2011). Students' perceptions of classroom instructional environments in the context of 'Universal design for learning.' *Learning Environments Research, 14*(2), 171-185. doi:10.1007/s10984-011-9090-2

Al-Azawei, A., Serenelli, F., & Lundqvist, K. (2016). Universal Design for Learning (UDL): A content analysis of peer reviewed journals from 2012 to 2015. *Journal of the Scholarship of Teaching and Learning 16*(3), 39–56 doi:https://doi.org/10.14434/josotl.v16i3.19295

AlRawi, J. M., & AlKahtani, M. A. (2021). Universal design for learning for educating students with intellectual disabilities: a systematic review. *International Journal of Developmental Disabilities*. doi:https://doi.org/10.1080/20473869.2021.1900505

Basham, J. D., Meyer, H., & Perry, E. (2010). The Design and Application of the Digital Backpack. *Journal of Research on Technology in Education, 42*(4), 339-359. doi:10.1080/15391523.2010.10782555

Booth, T., & Ainscow, M. (2002). *Index for Inclusion: Developing learning and participation in schools*: Centre for Studies on Inclusive Education.

Boyle, C., & Anderson, J. (2020). The justification for inclusive education in Australia. *Prospects*(49), 203–217. doi:https://doi.org/10.1007/s11125-020-09494-x

Bray, A., Devitt, A., Banks, J., Sanchez Fuentes, S., Sandoval, M., Riviou, K., Byrne, D., Flood, M., Reale, J., & Terrenzio, S. (2024). What next for Universal Design for Learning? A systematic literature review of technology in UDL implementations at second level. *British Journal of Educational Technology*, *55*(1), 113–138.

Capp, M. J. (2017). The effectiveness of universal design for learning: A meta-analysis of literature between 2013 and 2016. *International Journal of Inclusive Education, 21*(8), 791–807.

Capp, M. J. (2020). Teacher confidence to implement the principles, guidelines, and checkpoints of universal design for learning. *International Journal of Inclusive Education, 24*(7), 706-720. doi:10.1080/13603116.2018.1482014

CAST. (2022). Research Evidence. Retrieved from https://udlguidelines.cast.org/more/research-evidence#checkpoints

Chardin, M., & Novak, K. (2021). *Equity by Design: Delivering on the Power and Promise of UDL*. London: Sage Publications.

Daley, S. G., Hillaire, G., & Sutherland, L. M. (2016). Beyond performance data: Improving student help seeking by collecting and displaying influential data in an online middle-school science curriculum. *British Journal of Educational Technology, 47*(1), 121-134. doi:10.1111/bjet.12221

Daley, S. G., Xu, Y., Proctor, C. P., Rappolt-Schlichtmann, G., & Goldowsky, B. (2020). Behavioral Engagement among Adolescents with Reading Difficulties: The Role of Active Involvement in a Universally Designed Digital Literacy Platform. *Reading & Writing Quarterly, 36*(3), 278-295. doi:10.1080/10573569.2019.1635545

Dymond, S. K., Renzaglia, A., Rosenstein, A., Eul, J., Chun, Banks Ronald, A., . . . Gibson Christie, L. (2006). Using a Participatory Action Research Approach to Create a Universally Designed Inclusive High School Science Course: A Case Study. *Research & Practice for Persons with Severe Disabilities, 31*(4), 293-308. doi:10.1177/154079690603100403

Ender, K. E., Kinney, B. J., Penrod, W. M., Bauder, D. K., & Simmons, T. (2007). Achieving Systemic Change with Universal Design for Learning and Digital Content. *Assistive Technology Outcomes & Benefits, 4*(1), 115-129.

Fovet, F. (2019). Not just about disability: Getting traction for UDL implementation with International Students. In K. N. S. Bracken (Ed.), *Transforming Higher Education through Universal Design for Learning: An International Perspective*. London: Routledge.

Fovet, F. (2020). Universal Design for Learning as a Tool for Inclusion in the Higher Education Classroom: Tips for the Next Decade of Implementation. *Education Journal, 9*(6), 163-172.

Fovet, F. (2021a). Developing an Ecological Approach to the Strategic Implementation of UDL in Higher Education. *Journal of Education and Learning, 10*(4), 27-39.

Fovet, F. (2021b). UDL in Higher Education: a Global Overview of the Landscape and its Challenges. In F. Fovet (Ed.), Handbook of Research on Applying Universal Design for Learning Across Disciplines: Concepts, Case Studies, and Practical Implementation: IGI Global.

Galkienė, A., & Monkevičienė, O. (2021). The Model of UDL Implementation Enabling the Development of Inclusive Education in Different Educational Contexts: Conclusions. In A. Galkiene, Monkeviciene, O. (Ed.), *Improving Inclusive Education through Universal Design for Learning. Inclusive Learning and Educational Equity* (Vol. 5): Springer, Cham.

García, O., & Li, W. (2014). *Translanguaging: Language, Bilingualism and Education*. London: Palgrave Macmillan.

Gough, D. (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education, 22*(2), 213-228. doi:10.1080/02671520701296189

Grynova, I., & Kalinichenko, M. (2018). Trends in Inclusive Education in the USA and Canada. *Comparative Professional Pedagogy*(8), 2.

Higgins, J. P. T., & Greene, S. (Eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration. www.handbook.cochrane.org.

Hitchcock, C. H., Rao, K., Chang, C. C., & Yuen, J. W. L. (2016). TeenACE for Science: Using Multimedia Tools and Scaffolds to Support Writing. *Rural Special Education Quarterly, 35*(2), 10-23. doi:http://dx.doi.org/10.1177/875687051603500203

Katz, J. (2013). The Three Block Model of Universal Design for Learning (UDL): Engaging Students in Inclusive Education. *Canadian Journal of Education, 36*(1), 153-194.

Katz, J. (2015). Implementing the Three Block Model of Universal Design for Learning: effects on teachers' self-efficacy, stress, and job satisfaction in inclusive classrooms K-12. *International Journal of Inclusive Education, 19*(1), 1-20. doi:10.1080/13603116.2014.881569

Katz, J., & Sokal, L. (2016). Universal Design for Learning as a Bridge to Inclusion: A Qualitative Report of Student Voices. *International Journal of Whole Schooling, 12*(2), 36-63.

Kennedy, M. J., Thomas, C. N., Meyer, J. P., Alves, K. D., & Lloyd, J. W. (2014). Using Evidence-Based Multimedia to Improve Vocabulary Performance of Adolescents

With LD: A UDL Approach. *Learning Disability Quarterly, 37*(2), 71-86. doi:10.1177/0731948713507262

King-Sears, M. E., & Johnson, T. M. (2020). Universal Design for Learning Chemistry Instruction for Students With and Without Learning Disabilities. *Remedial & Special Education, 41*(4), 207-218. doi:10.1177/0741932519862608

King-Sears, M. E., Johnson, T. M., Berkeley, S., Weiss, M. P., Peters-Burton, E. E., Evmenova, A. S., . . . Hursh, J. C. (2015). An Exploratory Study of Universal Design for Teaching Chemistry to Students With and Without Disabilities. *Learning Disability Quarterly, 38*(2), 84-96. doi:10.1177/0731948714564575

Kortering, L. J., McClannon, T. W., & Braziel, P. M. (2008). Universal Design for Learning: A Look at What Algebra and Biology Students With and Without High Incidence Conditions Are Saying. *Remedial and Special Education, 29*(6), 352-363. doi:http://dx.doi.org/10.1177/0741932507314020

Lobascher, S. (2011). What are the potential impacts of high-stakes testing on literacy education in Australia? *Literacy Learning : the Middle Years, 19*(2), 9-19.

Loreman, T. (2017). Pedagogy for Inclusive Education. In *Oxford Research Encyclopedias*. London: Oxford Research Encyclopedias.

Lowrey, A., & Smith, S. (2018). Inclusion. Including Individuals With Disabilities in UDL Framework Implementation: Insights From Administrators 6(2), 127–142.

Mangiatordi, A., & Serenelli, F. (2013). Universal Design for Learning: A meta-analytic review of 80 abstracts from peer reviewed journals. *Research on Education and Media* (5), 109-118.

Marino, M. T. (2009). Understanding How Adolescents with Reading Difficulties Utilize Technology-Based Tools. *Exceptionality, 17*(2), 88-102. doi:10.1080/09362830902805848

Marino, M. T., Black, A. C., Hayes, M. T., & Beecher, C. C. (2010). An Analysis of Factors that Affect Struggling Readers' Achievement during a Technology-Enhanced STEM Astronomy Curriculum. *Journal of Special Education Technology, 25*(3), 35-47.

Marino, M. T., Gotch, C. M., Israel, M., Vasquez, E., Basham, J. D., & Becht, K. (2014). UDL in the Middle School Science Classroom: Can Video Games and Alternative Text Heighten Engagement and Learning for Students With Learning Disabilities? *Learning Disability Quarterly, 37*(2), 87-99. doi:10.1177/0731948713503963

McMahon, D., Wright, R., Cihak, D., Moore, T., & Lamb, R. (2016). Podcasts on Mobile Devices as a Read-Aloud Testing Accommodation in Middle School Science Assessment. *Journal of Science Education & Technology, 25*(2), 263-273. doi:10.1007/s10956-015-9591-3

Murphy, M. P. (2021). Belief without evidence? A policy research note on Universal Design for Learning. *Policy Futures in Education, 19*(1), 7-12. doi:doi:10.1177/1478210320940206

Newman, M., & Gough, D. (2020). Systematic Reviews in Educational Research: Methodology, Perspectives and Application: Springer VS, Wiesbaden.

Ok, M. W., Rao, K., Bryant, B. R., & McDougall, D. (2017). Universal Design for Learning in pre-K to grade 12 classrooms: A systematic review of research. *Exceptionality, 25*(2), 116–138. doi:https://doi.org/10.1080/09362835.2016.1196450

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2017). Overviews in Education Research: A Systematic Review and Analysis. *Review of Educational Research, 87*(1), 172-203. doi:doi:10.3102/0034654316631117

Rao, K., Smith, S. J., Edyburn, D., Grima-Farrell, C., Van Horn, G., & Yalom-Chamowitz, S. (2018). UDL Reporting Criteria. Developed by a working group of the Universal Design for Learning Implementation and Research (UDL-IRN) Research Committee. Retrieved from https://udl-irn.org/udl-reporting-criteria/

Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology, 95*(2), 375-392. doi:10.1037/0022-0663.95.2.375

Robinson, D. E. (2017). Universal Design for Learning and School Libraries: A Logical Partnership. *Knowledge Quest, 46*(1), 56-61.

Schuelka, M. (2018). Implementing Inclusive Education: DFID.

Scott, L., Saddler, S., Thoma, C. A., Bartholomew, C., Alder, N., & Tamura, R. (2011). Universal Design for Transition: A Single Subject Research Study on the Impact of UDT on Student Achievement, Engagement, and Interest. *Journal on Educational Psychology, 4*(4), 21-32.

Scott, L., & Temple, P. (2017). A Conceptual Framework for Building UDL in a Special Education Distance Education Course. *Journal of Educators Online, 14*.

Seok, S., DaCosta, B., & Hodges, R. (2018). A Systematic Review of Empirically Based Universal Design for Learning: Implementation and Effectiveness of Universal Design in Education for Students with and without Disabilities at the Postsecondary Level. *Open Journal of Social Sciences, 6*, 171-189. doi:10.4236/jss.2018.65014

Smith, S. J., & Harvey, E. E. (2014). K-12 online lesson alignment to the principles of Universal Design for Learning: the Khan Academy. *Open Learning, 29*(3), 222-242. doi:10.1080/02680513.2014.992402

Smith, S. J., Lowrey, K. A., Rowland, A. L., & Frey, B. (2020). Effective Technology Supported Writing Strategies for Learners With Disabilities. *Inclusion, 8*(1),

58-73. doi:10.1352/2326-6988-8.1.58

Staats, S., & Laster, L. A. (2018). Extending Universal Design for Learning through Concurrent Enrollment: Algebra Teachers' Perspectives. *Education Sciences, 8*, 1-19.

Su, Y.-L., & Reeve, J. (2011). A Meta-analysis of the Effectiveness of Intervention Programs Designed to Support Autonomy. *Educational Psychology Review, 23*(1), 159-188. doi:10.1007/s10648-010-9142-7

T&L National Forum. (2017). *Expanding our Understanding of Assessment and Feedback in Irish Higher Education*. Dublin, Ireland: National Forum for the Enhancement of Teaching and Learning in Higher Education Retrieved from https://hub.teachingandlearning.ie/resource/expanding-our-understanding-of-assessment-and-feedback-in-irish-higher-education/#

UNESCO. (2016). *Education 2030: Incheon Declaration and Framework for Action* United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000245656_eng

Verplaetse, L. S., & Schmitt, E. (2010). Bilingualism and Learning. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education (Third Edition)* (pp. 355-360). Oxford: Elsevier.

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (2019). *Systematic Reviews in Educational Research - Methodology, Perspectives and Application*: Springer.

Appendix 1. Review Article Summary Table

	Year	Authors	Setting	Sample	Participant Information (students)	Study Duration	Curriculum Area	Research Methodology	Research Instruments
1.	2011	Abell et al. (2011)	USA	867	Age range: 8 - 18 No. with low SES: not specified. *No. with AEN*: not specified. *Gender*: not specified *Ethnicity*: not specified	N/A	Diverse	Survey	Revised Individualised Classroom Environment Questionnaire (ICEQ-R)
2.	2010	Basham et al. (2010)	USA	35	*Age range*: Upper high school No. with low SES: not specified. No. with AEN: 1 *Gender*: 13 male, 22 female *Ethnicity*: majority African American	4-5 hours each cycle	Social Sciences	Design-Based Research (3 cycles)	All Cycles: Participant observations, video-recorded observations, field notes, student-generated artifacts. Cycles 2 & 3: student surveys and interviews
3.	2016	Daley et al. (2016)	USA	126	*Age range*: Grade 6 *No. with low SES*: School B: 63% No. with AEN: 12 *Gender*: 52 male, 74 female *Ethnicity*: *School A*: 82% white, 15% Asian American. *School B*: 98% African American	3-5 lessons	Science	Design-Based Research (1 cycle)	Electronic event usage log Written student inquiry questionnaires
4.	2020	Daley et al. (2020)	USA	315	*Age range*: Grades 6 - 8 No. with low SES: 78% *No. with AEN*: All, with 56% SEN *Gender*: 63% male, 37% female *Ethnicity*: Hispanic or Latino: 36%, White: 31%, Black or African	1 year (1 hour p week)	Literacy: Reading	Quantitative Quasi-experimental, pre-post-test design	Reading Comprehension pre/post tests Digital usage log in UDIO tool

					American: 26%, Asian or Pacific Islander: 2%, multi-racial: 4%				
5.	2006	Dymond et al. (2006)	USA	0	*Age range*: 9th Grade (mostly) No. with low SES: not specified. *No. with AEN*: Possible total of 33 (some may have attended more than one section or semester) *Gender*: not specified *Ethnicity*: White: 67.8%, Black: 24.2%, Asian/Pacific islander: 5.6%, Hispanic: 1.2%, Native American: 0.2%	1 year	Science	Participatory Action Research Case Study	Individual teacher pre and post intervention interviews; document analysis; meeting minutes and lesson plans; Post intervention teacher focus Group
6.	2007	Ender et al. (2007)	USA	0	*Age range*: not specified. No. with low SES: not specified. *No. with AEN*: not specified. *Gender*: not specified *Ethnicity*: not specified	3 years	Diverse	Not specified	Different evaluation instruments used in different schools. School level: Staff and Student survey and interviews; classroom observations; reports; school, teacher, and student products. Project Level: (a) Individualized Classroom Environment Questionnaire (ICEQ); (b) CATS scores and NCLB Adequate Yearly Progress (AYP) reports; and (c) monthly and year-end reports
7.	2016	Hitchcock et al.	USA	46	*Age range*: Grades 5 - 8	12 weeks	Literacy: Writing	Mixed methods:	Quant: pre/post writing tests:

		(2016)			No. with low SES: not specified. *No. with AEN*: Classroom A: 11%, Classroom B: 5% *Gender*: not specified *Ethnicity*: Hawaiian/part-Hawaiian: 70 - 85%			within person pre/post intervention comparison and naturalistic case study	Woodcock Johnson III, CBM Rubric Teacher and Student evaluations Qual: teacher/student surveys and focus groups
8.	2013	Katz (2013)	Canada	631	*Age range*: Grades 1 - 12 No. with low SES: not specified. *No. with AEN*: Students with severe learning disabilities, or very low proficiency in English were excluded. *Gender*: 51.5% male, 48.5% female *Ethnicity*: 87.5% born in Canada, 12.5% immigrated to Canada	¿?	Diverse	Quasi-experimental control Group pre/post-test design	observations, surveys, and Scales: Global Portrait of social and Moral Health for youth (GPDMHY) Acceptance of outgroups scale (CDP) Self-description Questionnaire (self-concept)
9.	2015	Katz (2015)	Canada	58 (teachers)	*Age range*: Grades 1 - 12 No. with low SES: not specified. *No. with AEN*: not specified. *Gender*: not specified *Ethnicity*: not specified	¿?	Diverse	Quasi-experimental control Group pre/post-test design	Interviews and Surveys (pre/mid/post)
10.	2016	Katz and Sokal (2016)	Canada	101	Age range: K-12 No. with low SES: not specified. No. with AEN: 11 *Gender*: 52 male, 49 female *Ethnicity*: 19 Aboriginal (First Nations, Metis, or Inuit)	¿?	Diverse	Qualitative analysis pre-post intervention.	Qualitative interviews

11.	2014	Kennedy et al. (2014)	USA	141	*Age range*: Grade 10 No. with low SES: 83% No. with AEN: 32 *Gender*: 76% male, 24% female *Ethnicity*: 63.3% African American, 26.7% Hispanic, 10% Caucasian	8 weeks	History	Quasi-experimental study	Pre-test and post test scores curriculum-based measures-probes of terms/historical figures and definitions
12.	2015	King-Sears et al. (2015)	USA	56	*Age range*: Grades 10 - 12 No. with low SES: 9 No. with AEN: 19 *Gender*: 25 male, 35 female *Ethnicity*: 27 white, 7 African American, 11 Hispanic, 11 Asian/Pacific Islander, 4 Multiracial	1 class (with 4-week delayed post-test)	Chemistry	Randomised Control Trial	Pre/Post and Delayed Post-test using Mole conversion tests Social Validity Questionnaire
13.	2020	King-Sears and Johnson (2020)	USA	44	*Age range*: High school No. with low SES: 21 *No. with AEN*: 16 (Study 1: 10; Study 2: 6) *Gender*: 24 male, 18 female *Ethnicity*: 5 White, 13 African American, 22 Hispanic, 2 Multiracial	2 sessions	Chemistry	Two quasi-experimental studies	Pre- and post-tests and social validity questionnaire
14.	2008	Kortering et al. (2008)	USA	290	*Age range*: High school No. with low SES: not specified. No. with AEN: 72 *Gender*: not specified *Ethnicity*: not	1 90-minute session	Maths and Biology	Survey	Student Self-Reported Engagement Scale Questionnaire plus some open questions

					specified				
15.	2009	Marino (2009)	USA	1153	*Age range*: Middle school: grades 6 - 8 No. with low SES: not specified. *No. with AEN*: 126 with severe reading difficulties, 205 poor readers *Gender*: 50% male, 50% female *Ethnicity*: 91% White, 5% Asian, 3% Hispanic, 1% African American	4 weeks	Science - astrophysics	Pre/Post-test Intervention Study (no control group)	1. Pre/post-test of scientific concepts, processes, and vocabulary (paper and pencil 25-item multiple-choice) 2. Six open-ended paper and pencil solutions forms 3. Degrees of Reading Power (DRP). 4. Tool Use Log. 5. Observations
16.	2010	Marino et al. (2010)	USA	1153	*Age range*: 10 - 14 No. with low SES: not specified. *No. with AEN*: ~ 11.5% (based on average across participants from the 3 schools) *Gender*: 50% male, 50% female *Ethnicity*: 91% White, 5% Asian, 3% Hispanic, 1% African American	4 weeks	Science - astrophysics	Pre/Post-test Intervention Study (no control group)	1. Pre/post-test of scientific concepts, processes, and vocabulary (paper and pencil 25-item multiple-choice) 2. Six open-ended paper and pencil solutions forms 3. Degrees of Reading Power (DRP). 4. Tool Use Log. 5. Observations
17.	2014	Marino et al. (2014)	USA	341	*Age range*: 10 - 14 No. with low SES: 56% No. with AEN: 57 *Gender*: 51% male, 49% female *Ethnicity*: 87% White	1 year	Science - pathogens	Mixed-methods design. ABAB intervention model	Paper-and-pencil pre-/posttest, data collected through video game play, student post-intervention focus group interviews
18.	2016	McMahon et al. (2016)	USA	47	*Age range*: Grade 6 No. with low SES: ~97.4% No. with AEN: 47 *Gender*: 57.4% male, 42.6% female		Science (assessment context)	Generalised Latin squares comparative study - differences within and between groups	Pre/Post control/experimental Test scores

					Ethnicity: 76.6% African American, 17% Caucasian, 4.3% Hispanic, 2.1% Native American				
19.	2017	Robinson (2017)	USA	5	*Age range*: Middle school No. with low SES: not specified. No. with AEN: 5 *Gender*: not specified *Ethnicity*: not specified	3 months	History (in library setting)	Case study	No explicit data collection
20.	2011	Scott et al. (2011)	USA	6	*Age range*: 15 - 19 No. with low SES: not specified. No. with AEN: 6 *Gender*: 4 male, 2 female *Ethnicity*: 2 Caucasian, 2 African American, 1 Hispanic/Latino, 1 Other		Social studies	ABAC multiple-treatment design	observation and survey, Student academic achievement (worksheet that included 10 questions)
21.	2020	Smith et al. (2020)	USA	730	*Age range*: Grade 6 No. with low SES: 46% No. with AEN: 75 *Gender*: 49% male, 51% female *Ethnicity*: 57% white, 11% African American, 25% Hispanic, and 7% "other"	1 year	Literacy: Writing		WRITE Progress Monitoring tool (WPM)
22.	2018	Staats and Laster (2018)	USA	31 (teachers)	Age range: N/A No. with low SES: N/A No. with AEN: N/A Gender: N/A Ethnicity: N/A		Maths	Qualitative. exploratory case study	Focus groups