Journal of Developmental and Physical Disabilities
https://doi.org/10.1007/s10882-023-09890-z

REVIEW ARTICLE

A Meta-analysis of Science Education Studies for Students with Intellectual and Developmental Disabilities (IDD)

Mehmet D. Sulu[1,2] · **Ronald C. Martella**[1] · **Orhan Aydin**[3] · **Virginia Bolshokova**[1] · **Emine Erden**[4]

Accepted: 9 January 2023

Abstract

Teaching science education has remained limited for students with intellectual and developmental disabilities (IDD), which, in turn, has resulted in an ongoing discrepancy between these students and their typically developing peers for decades. Although there is a growing body of research in effective teaching approaches aimed at overcoming this discrepancy, there is still a need to identify evidence-based practices for addressing this academic core content. The purpose of this meta-analysis was to (a) find out the skills taught in science education to students with IDD, (b) define the characteristics of instructional approaches or adaptations of instructional approaches used to teach science content and practices, (c) conduct visual and effect size analysis of science education studies meeting the Council for Exceptional Children (CEC) quality indicators (*QIs*; Cook et al., 2015), and (d) determine whether there are differences in effect sizes of science education studies meeting CEC *QIs* based on participant and intervention characteristics. Of 27 studies reviewed, 18 studies met all the CEC *QIs*. A meta-analysis of these 18 studies resulted in an overall medium effect size of 0.82 CI_{95} (0.76, 0.87). While all the moderator variables showed a medium effect size in participant characteristics, intervention characteristics showed differences in effect sizes for comprehension-based learning and peer and researcher-implemented interventions.

Keywords Science education · Students with IDD · Meta-analysis

✉ Mehmet D. Sulu
sulum1@nku.edu

1 College of Education, Purdue University, West Lafayette, USA

2 Educational Leadership and Advanced Studies, Northern Kentucky University, Highland Heights, USA

3 Department of Special Education, Anadolu University, Eskişehir, Türkiye

4 Faculty of Education, Hasan Kalyoncu University, Gaziantep, Türkiye

Published online: 09 February 2023

The importance of teaching scientific content knowledge, literacy, and thinking to all learners is stressed in legislation along with the Next Generation Science Standards (NGSS) (Apanasionok et al., 2019). Scientific content knowledge allows for the posing of questions and sharing discoveries (Spooner et al., 2011), where learners' skills to think, imagine, and continue to pursue rational and logical explanations for the observed universe are always encouraged (Mastropieri et al., 2006; Scruggs & Mastropieri, 2007). Acquiring scientific content and practices also have practical importance as learners need to use these thinking skills in real-life circumstances such as home, community, and school (Knight et al., 2020), and to make meaningful decisions about various issues such as ecological environment and well-being (Mastropieri & Scruggs, 1992; National Research Council, 2012; Spooner et al., 2011).

Such understanding is particularly important in the twenty-first century as political issues and moral dilemmas such as global warming can be confronted by our society (Osborne, 2007); thus, there is a need for increased emphasis on science education for all learners.

Although the importance of teaching the content and practices of science has been acknowledged by many researchers in the field (Spooner et al., 2017), science education for students with IDD has always been limited. This discipline includes a variety of complex academic vocabulary terms that requires prior knowledge of word meanings that are taught in very short time periods (Brigham et al., 2011; Ehsan et al., 2018). However, students often do not receive the adequate instruction needed to enhance their understanding and comprehension of science, nor do they develop the prerequisite background knowledge (Mason & Hedin, 2011), particularly for those with severe disabilities (Spooner et al., 2017). As a result, the academic achievement gap in science education between students with IDD and their typically developing peers is concerning (National Center for Educational Statistics, 2015), and this discrepancy appears to be greater than any other academic content area, such as literacy and math (Wei et al., 2015). Therefore, there is a need to close this gap and improve the scientific content and practices of students with IDD.

Two instructional models have been commonly used to teach science concepts in the literature—inquiry-based learning (e.g., Barthlow & Watson, 2014; Meij et al., 2015) and explicit and systematic instruction (e.g., Knight et al., 2018a). Of these two approaches, a growing body of research demonstrates that explicit and systematic instruction can be an effective approach to enhancing the scientific content knowledge of students with IDD (Apanasionok et al., 2019; Barnett et al., 2018; Spooner et al., 2011). According to findings from multiple meta-analyses, explicit and systematic instructional approaches can have large effects on science achievement with students with a wide range of disabilities, particularly in science vocabulary and science text reading (Kaldenberg et al., 2015; Therrien et al., 2011, 2014). Additionally, systematic and quality review studies have indicated that the use of explicit and systematic instruction to teach science concepts has become predominant in the field (Apanasionok et al., 2019), and these instructional approaches have greater efficacy than inquiry-based learning (Therrien et al., 2014).

Despite the effective and extensive usage of explicit and systematic instruction, the implementation of this instructional approach has been focused predominantly on science vocabulary and memorization versus more complex science content knowledge (Scruggs & Mastropieri, 2007). Thus, inquiry-based learning has been promoted in the literature to teach more science contents and practices (Therrien et al., 2011), because inquiry-based learning emphasizes real-world experiences and depth of learning above verbal memorization (Scruggs & Mastropieri, 2007). Additionally, inquiry-based learning has been combined and structured with explicit and systematic instruction effectively (e.g., Therrien et al., 2011), which is defined as guided instruction (Cawley, 1994; Therrien et al., 2014). Research indicates that both inquiry-based learning and guided instruction can also be effective in teaching science education to students with disabilities (Scruggs et al., 1998; Therrien et al., 2011).

Instructional approaches to teaching science education have been placed into the following three categories in previous reviews and a meta-analysis: explicit and systematic instruction, inquiry-based learning, and a combination of these two instructional approaches in general (e.g., Therrien et al., 2011; Therrien et al., 2014; Rizzo & Taylor, 2016). However, a recent comprehensive research review conducted by Apanasionok et al. (2019) labeled two studies (i.e., Carnahan & Williamson, 2013; Carnahan et al., 2016) as comprehension-based instruction. According to Browder and Spooner (2011), the goal of comprehension-based instruction is that students learn how to apply their knowledge of narrative texts to comprehend the aspects of expository texts. Although this instructional approach was found to be an evidence-based practice based on two studies, Apanasionok et al. (2019) indicated there is a need for further studies using this instructional approach to teach science practices.

Given the diversity of instructional programs, the heterogeneity among students with IDD, and the changing trend toward teaching science content and practices to students with disabilities in the twenty-first century, there is a need for a systematic investigation to assess the efficacy of science education studies for students with IDD. Although there have been multiple review studies conducted to investigate and evaluate the quality of these studies (e.g., Apanasionok et al., 2019; Knight et al., 2020), these review studies aimed to evaluate the scope and the quality of the science education studies. A meta-analysis is needed to examine the efficacy of interventions along with the differential effects of variables (David et al., 2022). Such an analysis can quantify the magnitude of the effects of science education interventions and determine if the effect was moderated by specific participant and intervention characteristics. Therefore, the current study will contribute to the literature with the aggregation and comparison of single-case experimental designs (SCEDs) science education studies for students with IDD. The specific research questions are:

1. What skills are taught in science education to students with IDD?
2. What were the characteristics of instructional approaches or adaptations of instructional approaches used to teach science content and practices to students with IDD?
3. What are the visual and effect size analysis of science education studies for students with IDD?

4. Are there differences in effect sizes of science education studies when classified based on students (i.e., age groups, disability) and intervention characteristics (i.e., independent variable, intervention agent)?

Method

The search procedures involved two phases (see Fig. 1). During the first phase, the first and fourth authors reviewed the literature from three different databases in March 2021, and the search was updated in August 2021. Databases included ERIC, Education Research Complete, and PsycINFO. The first and third authors used the whole and truncated versions of the following terms science, science education, Physics, Chemistry, Biology, plant, animal, human body, materials,

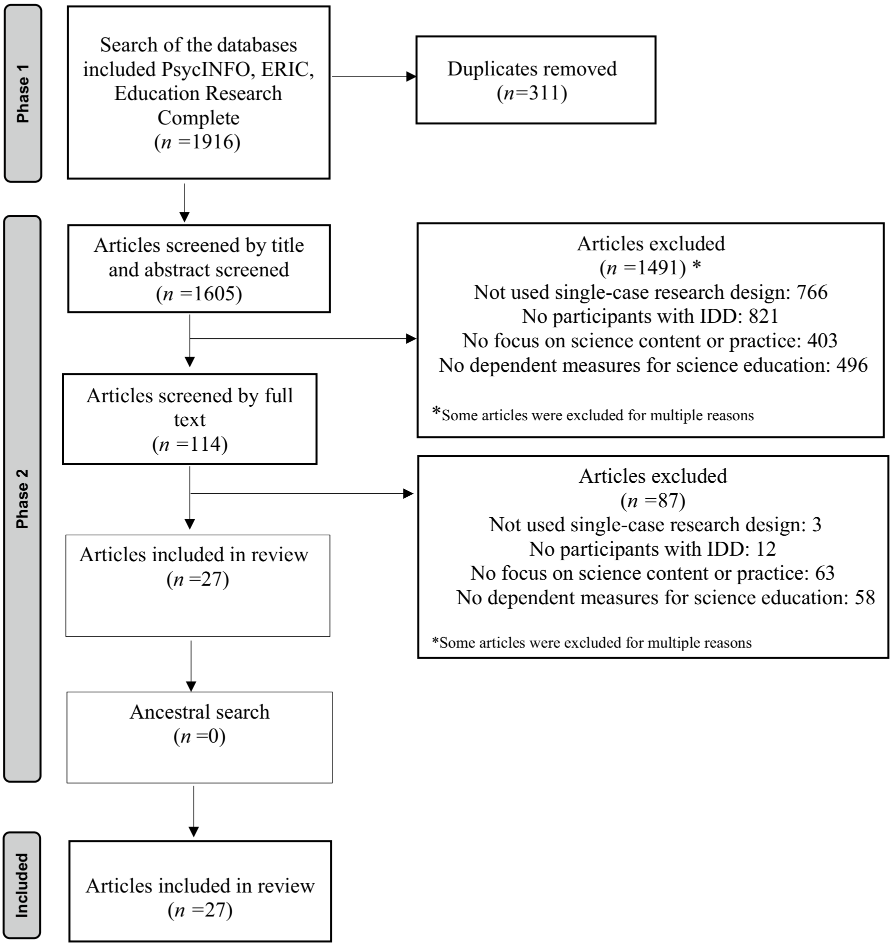

Fig. 1 PRISMA Flow Diagram of Literature Review

force, earth, electricity, acid, rocks, soil, magnets, space, chemical, weather, season, mass, planet, solar systems, living organisms, cells, fungus, insects, temperature, STEM, autism, ASD, autism spectrum disorder, intellectual disability, ID, mental retardation, developmental disability, down syndrome, pervasive developmental disorder, PDD, Asperger, intellectual deficiency, developmental impairment, handicap, neurodevelopmental disability. The authors limited the database search to those written in English and published in peer review journals. Additionally, studies that have been published in the last 10 years (2011–2021) were included to restrict this research to the most current studies due to the changing trend toward teaching science education (Knight et al., 2020; Osborne, 2014; NGSS Lead States, 2013). Although there is no consensus on the definition of "current" in the special education literature (Sulu et al., 2022), dynamic changes in science education studies have been observed in the last 10 years (e.g., comprehension-based learning, teaching science *content* and *practices*). Considering the limited scope of the science education studies in previous research (Knight et al., 2020), we only included studies published in the last 10 years.

After removing the duplicates ($n = 311$), the initial search yielded a total of 1605 studies. During the first phase, the first and fifth authors screened the abstracts and titles of the 1605 studies independently for inclusion. As a result of the first phase, 114 articles were selected for a full article review. Next, the first and fifth authors reviewed the titles, abstracts, and method sections of each of the 114 articles. Overall, 27 of the 114 studies met the inclusion criteria and were included in this review. Although an ancestral search of the included articles was conducted to ensure articles were not omitted in the electronic search, there were no other studies identified.

The authors reviewed the studies in these two phases based on the following inclusion criteria: (a) written in English, (b) published in a peer-review journal, (c) published between 2011 – 2021, (d) included at least one participant with IDD; however, if the participant was part of a group of several others who did not an IDD diagnosis, the data had to be disaggregated for the participant with IDD, (e) a focus on science content and/or practice, (f) dependent measures that assessed science content and practices, and (g) utilized SCEDs. IDD is defined in this study as having significantly below-average intellectual functioning and adaptive behavior problems. Example of disability labels that could be categorized under or with IDD for purposes of this review could include autism spectrum disorder (ASD), Down syndrome, Developmental Delay, ADHD, and other health impairments. If a study included participants with disabilities but without being labeled as IDD, the authors extracted and evaluated data for only those participants with an IDD diagnosis. Finally, this analysis included studies that used SCEDs because researchers in the field of special education utilize SCEDs predominantly compared to quasi and group experimental studies (Kennedy, 2005; Maggin et al., 2018). Additionally, SCEDs make it possible to draw more precise conclusions regarding the nature of behavior change that is distinctive to an individual. (Martella et al., 2013; Maggin et al., 2018).

Article Analyses

Several analyses were conducted for the review. These analyses included a quality appraisal, descriptive analysis, visual analysis, effect size calculations, and moderator variable identification.

Quality Appraisal

Articles were analyzed based on the 2014 CEC Standards for Evidence-Based Practices in Special Education (Cook et al., 2015). These standards included eight categories including (*QI-1*) *Context and setting*, (*QI-2*) *Participants*, (*QI-3*) *Intervention agent*, (*QI-4*) *Description of practice*, (QI-5) *Implementation fidelity*, (*QI-6*) Internal validity, (*QI-7*) *Outcome measures/Dependent variables*, and (*QI-8*) *Data analysis* (see Table 1).

Descriptive Analysis

The studies met all the *QIs* were coded descriptively. The categories included (a) participants, (b) setting(s), (c) practitioners, (d) research design, (e) dependent variable(s), (f) dependent measure(s), (g) instructional program, and (h) maintenance and generalization.

Visual Analysis

Visual analysis was conducted for the studies that met all *QIs*. The U.S. Institute of Education Sciences proposed six data aspects of SCEDs: level, trend, variability, overlap, immediacy of effect, and consistency of data patterns (What Works Clearinghouse-WWC, 2022). In this context, authors analyzed mean level change, trend direction, data variability, percentage of overlap, the immediacy of effect, and consistency of data patterns of single-case graphs (cf. Lane & Gast, 2014; WWC, 2022).

Effect Size Analysis

For effect size calculations of SCEDs graphs derived from the included studies were digitized. The PlotDigitizer, a reliable and valid software program for digitizing single-case graphical data (Aydin & Yassikaya, 2022), was used for data extraction.

The authors calculated Tau-*U* (Parker et al., 2011) values for the studies that met all CEC *QI*s. All Tau-*U* values were calculated via the web-based calculation engine http://singlecaseresearch.org/calculators/tau-u (Vannest et al., 2016). As suggested Brossart et al. (2018), the baseline correction box for A-B contrast was clicked on when A (baseline) contrast was ≥ .40.

Weighted Tau-*U* with the 95% CI (CI_{95}) for individual studies and moderating variables were calculated in the current study. The authors also calculated

Table 1 CEC Quality Indicator Definitions for Single-Case Designs

CEC Quality Indicators
QI-1.0 Context and Setting:
To meet QI-1.1. Studies needed to provide adequate information on demographic variables (e.g., age, region, numbers of participants) to identify context/setting.
QI-2.0 Participants:
This category included wo subcategories; QI-2.1 in which a study needed to describe at least one demographic variable of participants (i.e., sex, age/grade, and language status), and QI-2.2 where a study needed to provide sufficient information about diagnosing determination status of students with disabilities.
QI-3.0 Intervention Agent:
Top meet QI-3.1, studies needed to provide sufficient information on demographic background of intervention agent along with their role. IQ-3.2. is met when studies included information study had to include information regarding instructor implementing the science intervention training. Studies did not include sufficient information with respect to instructor' training was scored not meeting QI-3.2.
QI-4.0 Description of Practice:
To meet QI -4.1 a study had to describe the intervention process elaborately to allow for replication. Similarly, to meet QI-4.2 required a study to describe materials if the study used any. Studies did not give sufficient information regarding intervention process and materials were scored as not meeting the criteria for corresponding QI.
QI-5.0 Implementation Fidelity:
To meet QI-5.1 a study must have included fidelity assessment of intervention process that also required direct and reliable observations. To meet QI-5.2 a study had to involve sufficient information with respect to the dosage of interventions such as the length of intervention. To meet QI-5.3 a study must have supplied timing and frequency of fidelity assessment procedure by assessing adherence, and dosage of implementation process. A study had to replicate this process twice to meet QI-5.3.
QI-6.0 Internal Validity:
A study had to control independent variable systematically to meet QI-6.1, describe control/comparison conditions elaborately to meet QI-6.2, limit comparison-condition participants to access to the intervention firmly to meet QI 6.3, demonstrate three data points in three different phases to meet QI-6.5, include minimum 3 data points for all phases for single subject design studies QI-6.6, control common threats for internal validity QI-6.7.
QI-7.0 Outcome Measures/Dependent Variable
This category consists of six items. A study must have measured the socially validity to evaluate if the intervention was socially important to meet QI-7.1, clearly describe dependent variable measurement process to meet QI-7.2, report all important outcomes including positive and negative effects of intervention to meet QI-7.3, use appropriate timing and frequency implementation process to meet QI-7.4, measure interobserver agreement, and test–retest reliability at acceptable levels to meet QI-7.5.
QI-8.0 Data Analysis
To meet QI-8.2., had to include a clear single-subject design table depicting the results clearly.

aggregated Tau-*U* with CI_{95} to determine the overall effect for the included science education studies. Based on these results, the forest plots were used to show how Tau-*U* values were distributed over the CI_{95} range. A forest plot allows reporting the results of studies individually and combined. In addition, a forest plot is an indicator of the size and precision of the effect (Cooper et al., 2019).

Moderator Variables Analysis

The moderator variables were coded for each study meeting all CEC *QI*s. These moderators were defined and analyzed for similar features so that the results may lead to the shaping of policies and facilitate the national and international spread of effective practices (Cooper et al., 2019). Moderator analyses for single-case meta-analysis studies may help uncover the boundaries of interventions, including school levels, disability categories, and intervention procedures separately (Dowdy et al., 2021; Ledford & Gast, 2018). Moderators were determined according to the salient variables in the studies. In this context, we placed our moderators into four categories: (a) independent variables, (b) school levels, (c) disabilities, and (d) interventionists. Similar to Apanasionok et al. (2019), we categorized the independent variables in three subcategories— self-directed learning, explicit and systematic-instructional procedures, and comprehension-based learning. School levels were elementary (6 to11 years old), secondary (11 to 18 years old), and post-secondary (18 years old or older). Disabilities consisted of ASD, Intellectual Disability (ID), and both ASD and ID (ASD + ID). Interventionists included teachers, researchers, teachers and researchers, peers and researchers, facilitators and self-monitoring programs, and online applications.

Interrater Agreement

The reliability data were collected by the first, third, and fifth authors. The first author was a doctoral candidate in special education and a Board-Certified Behavior Analyst (BCBA), the third author was a doctoral candidate in special education, and the fifth author was an assistant professor in special education. The interrater agreement calculated through the number of agreements across raters was divided by the number of disagreements and number of agreements and multiplied by 100% across all the calculations. Disagreements were discussed between the disagreeing authors until they reached a consensus.

The interrater agreement was calculated in the following stages in the current study. First, while the first author reviewed the 1605 studies, the fifth author reviewed 321 (20%) studies during the search procedure. In the second phase, the first and the fifth authors reviewed the 114 studies inclusion. Agreement among the authors on the inclusion of the studies in both phases was 100%. Second, 27 studies were coded by the first author; 13 (48%) studies were randomly selected and coded by the fifth author for the CEC *QI*s. The interrater agreement was calculated as 94.7%. Next, descriptive analysis and moderator variables for the 18 studies that met the CEC *QIs* were coded by the first and the third author independently. The interrater agreement was 98.3% and 98.4%, respectively. Next, the third author conducted the data extraction and effect size calculations of the 18 studies, and the first author conducted 4 (22%) randomly selected studies for reliability. The interrater agreement was 100%. Lastly, the third author conducted a visual analysis of the 18 studies, and the first author did 11 (61%) randomly chosen studies for reliability. The interrater agreement was 100%.

Results

Quality Appraisal

Of the 27 studies included in this review, 18 studies (64%) met all the *QIs* outlined by Cook et al. (2015). Table 2 shows the coding for the 9 (36%) studies that did not meet the CEC standards.

Descriptive Analysis

Figure 2 depicts (a) participants (i.e., gender, school age, diagnosis), (b) setting(s), (c) practitioners, and (d) research design; and Table 3 shows the results for (e) dependent variable(s), (f) dependent measure(s), (g) instructional program, and (h) maintenance and generalization for the 18 studies that met all the CEC *QIs*.

Participants

A total of 64 participants with IDD were included across the 18 studies. The categories included the following: 43 (70%) students were diagnosed with ID, 14 (16%) students were diagnosed with ASD, and 7 (14%) students were diagnosed with ID and ASD. A total of 43 (67%) males and 21 (33%) females participated in 18 studies. Among 7 (39%) studies reporting ethnicities, 12 (19%) students were White, 4 (7%) students were Black, 4 (7%) students were Latinx, 1 student was Asian/Pacific Islander, 1 (2%) student was biracial, and 1 (2%) student was Native Hawaiian/Other Pacific Islander. IQ scores were reported in 12(67%) studies and ranged from 40 (Riggs et al., 2013) to 85 (McMahon et al., 2016).

Setting(s)

There was a range of study settings as the following: 4 (22%) studies were conducted in special education classrooms (nonspecified), 4 (22%) studies were conducted in general education and special education classrooms, 4 (22%) studies were conducted in special education resource rooms, 2 (11%) studies were conducted in special education self-contained classrooms, 1 (6%) study was conducted in general education classrooms, 1 (6%) study was conducted in a room adjacent to special education classroom, 1 (6%) study was conducted in a computer lab on a university campus, and 1 (6%) study was conducted in a private school for students with disabilities.

Practitioners

Of the 18 studies, the interventions were delivered by teachers in 10 (54%) studies, by researchers in 4 (22%) studies, by teachers and researchers in 1 (6%) study, by study peers and researchers in 1(6%) study, by facilitators and self-monitoring in 1 (6%) study, and by online applications in 1 (6%) study.

Table 2 Studies Not Meeting CEC QIs

	QI-1	QI-2		QI-3		QI-4		QI-5			QI-6									QI-7						Q-8		
	.1	.1	.2	.1	.2	.1	.2	.1	.2	.3	.1	.2	.3	.4	.5	.6	.7	.8	.9	.1	.2	.3	.4	.5	.6	.1	.2	.3
Collins et al. (2017)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	—	Y	Y	Y	—	—	Y	Y	Y	Y	Y	—	—	Y	—
Fetko et al. (2013)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	—	Y	Y	Y	—	—	Y	Y	Y	Y	Y	—	—	Y	—
Jimenez et al. (2012)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	—	Y	Y	Y	—	—	Y	Y	Y	Y	Y	—	—	Y	—
Jimenez et al. (2014)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	—	Y	Y	N	—	—	Y	Y	Y	Y	Y	—	—	Y	—
Knight et al. (2012)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	—	Y	Y	Y	—	—	Y	Y	Y	N	Y	—	—	Y	—
Knight et al. (2015)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	—	Y	Y	Y	—	—	Y	Y	Y	N	Y	—	—	Y	—
McKiss-ick et al. (2018)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	—	Y	Y	Y	—	—	Y	Y	Y	Y	Y	—	—	Y	—
Smith et. al. (2013b)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	—	Y	Y	Y	—	—	Y	Y	Y	Y	Y	—	—	Y	—
Starling et al. (2015)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	—	Y	N	N	—	—	Y	N	Y	N	Y	—	—	Y	—

Table 3 Descriptive Analysis

Article Citation	Dependent Variable(s)	Dependent Measure(s)	Instructional Program	Maintenance & Generalization
Carnahan and Williamson (2013)	Compare-contrast comprehending expository science text	Comprehension questions for passage and preposition scores on the Venn diagram.	GO and Venn diagram	LM: 6 weeks 2 M data points
Carnahan et al. (2016)	Expository science text comprehension.	Comprehension questions for each passage (i.e., 5 explicit, 5 inferential).	Text structure intervention (i.e., organization and summary sheet)	LM: 2 weeks 2 M data points
Collins et al. (2011)	Periodic table: gas, liquid, and solid.	Functional application in task-analyzed cooking activities	CTD	LM: Fallowing mastery criterion MM data points MG data points (i.e., baseline, intervention)
Greene and Bethune (2019)	Science vocabulary and core content in three units (i.e., energy, water, and plant anatomy).	Science vocabulary identification questions; verbal answers to unit concept questions	Group systematic instruction using errorless prompting procedures	LM: Immediately after in energy and weather units MM data points
Heinrich et al. (2016)	Science vocabulary (homeostasis and Punnett square).	Correct responses in chained and discrete tasks	Embedded simultaneous prompting procedure	LM: Immediately after MM data points 2 G data points after intervention
Hudson et al. (2014)	Science content comprehension via listening.	Wh- word open ended comprehension questions	System of least prompts intervention and read-aloud of adapted science lessons and self-monitoring	MG data points (i.e., baseline, after intervention)
Karl et al. (2013)	Science: describing the effects of forces on the effects of motions.	Science worksheets (i.e., 3 multiple choice and 2 True–False questions).	Simultaneous prompting procedure.	LM: 1,3-, and 5-weeks fallowing mastery criterion MM data points 2 G data points (i.e., baseline, after intervention)
Knight et al. (2018a)	Science units (i.e., five senses, rock cycle, earth and sky, and life cycle) from *Early Science Curriculum.*	6-item curriculum assessment test included 24 questions	Scripted lessons vs. unscripted task analyzed lessons	LM: Immediately after MM data points

Table 3 (continued)

Article Citation	Dependent Variable(s)	Dependent Measure(s)	Instructional Program	Maintenance & Generalization
Knight et al. (2018b)	Science vocabulary, comprehension, and application questions.	Science comprehension quizzes (i.e., 3 vocabulary, 3-literal comprehension, and 1 application question).	Supported eText using modified *Book Builder* included explicit instruction	N/A
Knight et al. (2013)	Science vocabulary comprehension (e.g., precipitation, condensation).	Correct responses in task analysis.	Systematic instruction (i.e., CTD, GO)	LM: 1 week 1 M data point for P1 and P2 MG data points (i.e., baseline, intervention, M)
McMahon et al. (2016)	Science vocabulary word lists: bones, organs, and plant cells.	20 questions (i.e., 10 multiple-choice, 10 labelling)	Augmented reality included images/objectives with user created digital content.	N/A
Miller and Taber-Doughty (2014)	Inquiry problem-solving activities (not specified)	Percent of task analysis steps completed independently	Self-monitoring checklists and five science guided inquiry lessons (i.e., question, observe, plan, experiment, and explain)	2 G data points (i.e., after intervention)
Miller et al. (2015)	Inquiry skills (i.e., Question, Observe, Plan, Experiment, and Explain)	Independent compliance of inquiry-based problem-solving steps	Guided science inquiry and self-monitoring checklist	LM: 2 weeks 1 M data point 6 G data points (i.e., baseline, intervention, M)
Riggs et al. (2013)	*Targeted:* Heredity *Nontargeted:* Transmission of genetic information *Anecdotal:* Science vocabulary related to meiosis	*Targeted:* Percentage of correct responses for dominant/ recessive traits and Punnett Square *Nontargeted:* Open ended questions based on scenario *Anecdotal:* Open ended questions	CTD and multiple exemplars training	LM: 1 to 3 weeks following mastery criterion 2 M data points 2 G data points (i.e., baseline, after intervention)
Rivera et al. (2017)	Science vocabulary (i.e., roots, stem, seeds, soil).	Science vocabulary identification questions Percent of task analysis steps completed independently.	Multicomponent multimedia shared story via an iPad included scripted instructions.	LM: Following mastery criterion 2 M data points 2 G data points (i.e., after intervention)

Table 3 (continued)

Article Citation	Dependent Variable(s)	Dependent Measure(s)	Instructional Program	Maintenance & Generalization
Roberts et al. (2020)	Science text comprehension in an expository science text. Body Fuel: A Guide for Good Nutrition Body Fuel: A Guide to Good Nutrition, and Teen Smoking: Current Issues	10 reading comprehension questions (i.e., 5 multiple-choice, 5 open-ended).	Multi-component literacy intervention (i.e., pre-teaching vocabulary, GO, summarizing)	LM: 3 weeks A 32-question summative assessment
Smith et al. (2013a)	Science units (i.e., five senses, rock cycle, earth and sky, and life cycle) from *Early Science Curriculum.*	Number of correct responses in unit assessments	Systematic instruction (i.e., inquiry, task analysis, CTD, examples/non-examples, KWHL)	LM: 5-week following mastery criterion 1 M data point for Units 1, 2, and 3
Wood et al. (2020)	Generating and answering science questions.	Researcher developed 5 questions, generating questions using and without using GO.	Technology based supports and systematic instruction included CTD	LM: 2 weeks 1–2 M data points MG data points (i.e., baseline, intervention, maintenance)

CTD Constant Time Delay, *G* Generalization, *GO* Graphic Organizers, *LM* Latency to Maintenance, *KWHL* What do I *K*now? *W*hat do I want to Know? *H*ow can I find out? What did I Learn?, *M* Maintenance, *MG* Multiple Generalization, *MM* Multiple Maintenance, *P1* Participant 1, *P2* Participant 2

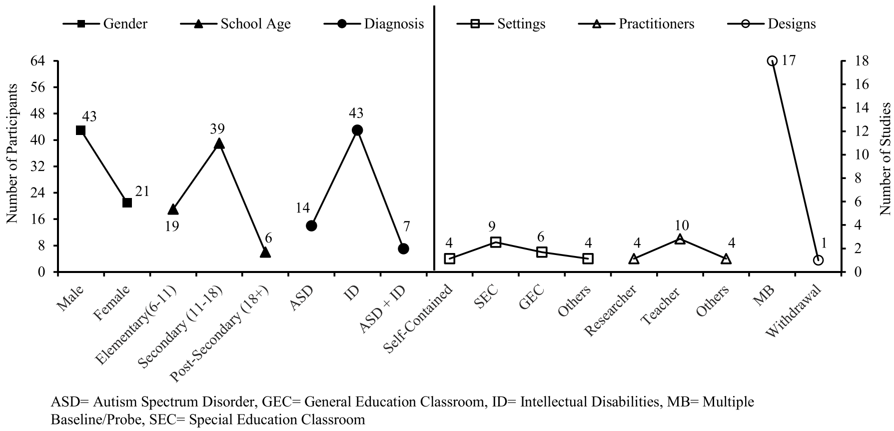

Fig. 2 Demographic information derived from 18 articles. ASD = Autism Spectrum Disorder, GEC = General Education Classroom, ID = Intellectual Disabilities, MB = Multiple Baseline/Probe, SEC = Special Education Classroom

Research Design and Finding(s)

Of the 18 studies, all but one (i.e., Carnahan & Williamson, 2013—withdrawal design), used a multiple probe design.

Dependent Variable(s)

Of the 18 studies, the dependent variable was science vocabulary in 6 (33%) studies such as energy, plant anatomy, homeostasis, precipitation, condensation, bones, organs, plant cells, roots, seeds, and mitosis. The dependent variable was science concepts in 4 (22%) studies, including descriptions of the effects of forces, periodic table, and four science units (i.e., five senses, rock cycle, earth and sky, life circle). Three (16%) studies defined the dependent variable as science text comprehension skills, including compare-contrast comprehending expository science text, expository science text comprehension, and science text comprehension. Science inquiry skills were dependent variables in 3 (16%) studies, including inquiry problem-solving skills and inquiry problem-solving skills. One (6%) study investigated listening science content comprehension, and 1 (6%) study targeted generating and answering science questions.

Dependent Measure(s)

The primary dependent measure across the majority of studies was the number of correct answers on daily implemented probe sessions with a total of 12 (66%) studies. Other studies used varying methods of determining performance. These methods included the following: 2 (11%) studies utilized the completion of a task analysis of digital assessment that included nine steps, 2 (11%) study curriculum-based probe assessments that were included in the *Early Science Curriculum*, 1 (6%) study used the completion of a task analysis of the functional assessment of the periodic table via cooking activities, 1 (6%) study used correct science vocabulary responses in 16-step task analysis, and 1 (6%) study used inquiry problem-solving skills in a 5-step self-monitoring checklist.

Additionally, the same questions were provided in a different order in 3 (16%) studies across baseline and intervention conditions (i.e., multiple-choice questions to assess science vocabulary and content which was administered across 30 probe sessions [$n=1$], worksheet questions that included multiple choice and true false questions to assess descriptions of the effects of forces on the effects of motions [$n=1$], and slideshows to assess science vocabulary across probe sessions [$n=1$]).

Instructional Program

A wide variety of instructional approaches were used across the 18 studies. Thirteen studies (72%) used explicit and systematic instruction that included prompting procedures, technology-based supports, graphic organizers, and an *Early Science Curriculum*. Of the 5 remaining studies, 3 (16%) utilized self-directed learning, and 2 (11%) employed comprehension-based learning.

Maintenance and Generalization

Maintenance data were collected across 14 (77%) studies. Latency to maintenance (i.e., the elapsed time between the final intervention data point and the first maintenance data point) varied across studies. Four (22%) studies measured maintenance data immediately after the completion of the intervention or after participants met a mastery criterion. The latency to maintenance was between 1 and 6 weeks in 10 (55%) studies.

Of these 14 studies, the number of maintenance data points ranged from 1 to 17. Only 3 (16%) studies collected multiple maintenance (i.e., 3 or more data points for each participant) data points. Three (16%) studies did not collect maintenance data across all participants or behaviors. Maintenance data results appeared positive for 12 (66%) of the 14 studies, and 2(11%) studies had mixed results.

Generalization data were collected across 10 (55%) of the 18 studies. Of the 10, 3 (16%) provided two generalization data points after the intervention condition, 1 (6%) provided two generalization data points (however, one of the data points was before the intervention condition and one was after the intervention condition), and the remaining 5 (27%) studies included a minimum of three data points. Of these 5 studies, generalization data were collected during baseline, intervention, and maintenance conditions. Of the 11 studies that assessed generalization, findings appeared to be positive in 9 (50%) studies; however, there was a lack of generalization of science outcomes to untrained stimuli in 2 (11%) studies.

Visual Analysis

As shown in Table 4, the average of the level change and the trend direction showed differences based on the representation of data units in each study. There were accelerating trend directions and consistent data patterns in all the conditions of 16 (89%) studies. Only one condition had decelerating trend direction and no consistent data patterns in 2 (11%) studies; however, the remaining conditions had accelerating trend directions and consistent data patterns in both. While 3 (16%) studies had 100% data stability across all the conditions, data stability and variability for each condition showed differences in the 15 (84%) studies. The average percentage of overlap was 0% for 3 (16%) studies, and the remaining studies had varying degrees of overlapped data.

Effect Size Analysis

Tau-*U* (Parker et al., 2011) effect size values were calculated for the science education studies that met CEC *QI*s and for the predetermined moderators. Before measuring the effect sizes, raw data were derived from the single-case graphs presented in the studies through the data extraction software (i.e., PlotDigitizer).

In total, 141 AB graphs were derived from 18 articles. A total of 2158 raw data points were extracted from 141 graphs via the PlotDigitizer. Each of the raw data was verified by visually analyzing the graphs and rounding to the nearest integer value, as suggested by Aydin and Yassikaya (2022). Approximately 40% (872 data points, mean = 6.18, range = 3 to 24) of this raw data were from the baseline (A) condition, whereas 60% (1286 data points, mean = 9.12, range = 3 to 50) were from the intervention (B) condition.

Table 5 shows weighted Tau-*U* effect sizes and other related information of the 18 articles. In addition, a forest plot in Fig. 3 shows weighted Tau-*U* values with CI_{95} for individual studies and the aggregated weighted Tau-*U* value with CI_{95} for the overall effect. According to the results based on the 18 individual studies, 9 studies had a medium effect, six had a strong effect, and three had a small effect.

Overall, Tau-*U* calculations were performed with 8484 data pairs from the 141 AB graphs. Tau-*U* with baseline correction (Tau-U_{BC}) was performed due to A contrast ≥ .40 for 26 of the 141 AB graphs; Tau-*U* calculations (non-trend) were performed for 115 AB graphs. The aggregated weighted Tau-*U* for all studies was .820 CI_{95} (.762, .879). The overall effect size of science education studies was medium. It is important to note that the confidence interval range in effect sizes for individual studies had large range. This, the individual results of science education studies have potential to positively impact student outcomes, albeit additional factors may need to be considered for these effects. The confidence interval range for the overall effect size for each individual study was small. That is, this result is a more precise and credible indication that science education studies have a medium effect.

Table 4 Visual Analysis of 18 Articles

Studies	Level Change (Average)	Imme-diacy of Effect (Average)	Trend Direction (Each Condition)	Stability/ Variability (Each Condition)	Percentage of Overlap (Average)	Consistency of Data Patterns
Carnahan and Williamson (2013)	30.8	30.5	Accelerating	100% S (all Cs)	8.8%	Yes
Carnahan et al. (2016)	42.3	37.6	Accelerating	100% S (all Cs)	0%	Yes
Collins et al. (2011)	28.2	14	Accelerating (5 Cs) Decelerating (1 C)	86% S-14% V (for 2); 36% S- 64% V (for 1); 98% S-2% V (for 1); 51% s-49% V (for 1); 100% V (for 1)	65%	Yes (5 Cs) No (1 C)
Greene and Bethune (2019)	2.4	.8	Accelerating (8 Cs) Decelerating (1 C)	89% S-11% V (for 2); 22% S-78% V (for 1); 100% S (for 2); 56% S-44% V (for 1); 63% S-37% V (for 2); 67% S-33% V (for 1)	48%	Yes (8 Cs) No (1 C)
Heinrich et al. (2016)	61.5	30	Accelerating	60% S-40% V (for 1); 43% S- 57% V (for 1)	27%	Yes
Hudson et al. (2014)	1.8	1.1	Accelerating	100% S (for 1); 33% S-67% V (for 1); 47% S-53% V (for 1)	51%	Yes
Karl et al. (2013)	70.2	28.7	Accelerating	28% S-72% V (for 1); 57% S-43% V (for 1); 60% S-40% V (for 1); 30% S-70% V (for 1)	6%	Yes
Knight et al. (2013)	8.5	5.3	Accelerating	57% S-43% V (for 1); 38% S-62% V (for 1); 71% S-29% V (for 1)	13%	Yes
Knight et al. (2018a)	2.3	1.5	Accelerating	100% S (for 7); 75% S-25% V (for 2); 50% S-50% V (2); 38% S-62% V (2); 60% S-40% V (2); 44% S-56% V; 67% S-33% V (2); 58% S-42% V (2); 14% S-86% V (for 1); 77% S-23% V (1); 42% S-58% V (1); 63% S-37% V (1); 56% S-44% V (1); 94% S-6% V (1); 62% S-38% V (1); 70% S-30% V (1); 31% S-69% V (1); 22% S-78% V (1); 78% S-22% V (1)	26%	Yes
Knight et al. (2018b)	4.2	2.3	Accelerating	70% S-30% V (for 1); 45% S-55% V (for 1); 100 S (for 1); 88 S-%12 V (for 1)	17%	Yes

Table 4 (continued)

Studies	Level Change (Average)	Imme-diacy of Effect (Average)	Trend Direction (Each Condition)	Stability/ Variability (Each Condition)	Percentage of Overlap (Average)	Consistency of Data Patterns
McMahon et al. (2016)	5.6	4	Accelerating	60% S- 40% V (for 3); 43% S- 57% V (for 3); 80% S- 20% V (for 3); 50% S- 50% V (for 3); 29% S- 71% V (for 2); 33% S- 67% V (for 1); 83% S- 17% V (for 1); 30% S- 70% V (for 1); 71% S- 29% V (for 1); 55% s- 45% V (for 1); 91% S- 09% V (for 1); 27% S- 73% V (for 1); 36% S- 64% V (for 1); 75% S- 25% V (for 1); 100% S (for 1);	7%	Yes
Miller and Taber-Doughty (2014)	92	93.4	Accelerating	100% S (all Cs)	0%	Yes
Miller et al. (2015)	42.2	46.5	Accelerating	100% S (for 3); 100% V (for 1); 50% s- 50% V (for 1); 75% S- 25% V (for 1)	8%	Yes
Riggs et al. (2013)	62	30.6	Accelerating	75% S- 25% V (for 2); 83% S- 17% V (for 1); 44% S- 56% V (for 1); 93% S- 7% V (for 1); 46% S- 54% V (for 1); 92% S- 8% V (for 1); 94% S- 6% V (for 1); 91% S- 9% V (for 1)	30%	Yes
Rivera et al. (2017)	70.6	53.6	Accelerating	100% S (for 1); 83% S- 17% V (for 1); 30% S- 70% V (for 1)	0%	Yes
Roberts et al. (2020)	43.5	34.4	Accelerating	100% S (for 1); 86% S- 14% V (for 1); 88% S- 12% V (for 1)	7%	Yes
Smith et al. (2013a, b)	4	2.6	Accelerating	63% S- 37% V (for 3); 57% S- 43% V (for 3); 60% S- 40% V (for 1); 33% S- 67% V (for 1); 80% S- 20% V (for 1); 55% S- 45% V (for 1); 40% S- 60% V (for 1); 83% S- 17% V (for 1)	21%	Yes
Wood et al. (2020)	4	3.4	Accelerating	63% S- 37% V (for 2); 100% S (for 2); 88% S- 12% V (for 2); 95% S- 5% V (for 1); 94% S- 6% V (for 1); 43% S- 57% V (for 1)	33%	Yes

C Condition, *S* Stability, *V* Variability

Moderating Variables Analysis

Table 6 shows weighted Tau-*U* effect sizes and the CI_{95} for moderator information. In addition, a forest plot in Fig. 4 shows weighted Tau-*U* values with the CI_{95} for the moderating variables across the 18 studies.

Independent Variables

Three independent variables were categorized—explicit and systematic instruction, self-directed learning, and comprehension-based learning. The 13 studies using explicit and systematic instructional procedures had the effect size of .789 CI_{95} (.724, .854) weighted Tau-*U*. On the other hand, the 3 self-directed learning studies had the effect size of .900 CI_{95} (.770, 1.00) weighted Tau-*U*. And, the 2 comprehension-based learning studies had the effect size of .938 CI_{95} (.647, 1.00) weighted Tau-*U*. While the most used independent variable in the studies was explicit and systematic instructional procedures (99 AB contrasts), the least used independent variable was comprehension-based learning (9 AB contrasts). While explicit and systematic instructional procedures and self-directed learning had a medium effect and a small CI_{95} range, comprehension-based learning had a strong effect and a large CI_{95} range (see Fig. 2).

School Levels

Grades were categorized into three levels—elementary, secondary, and post-secondary. The weighted Tau-*U* values of science education interventions were .764 CI_{95} (.599, .882) for elementary students .838, CI_{95} (.739, .937) for secondary students, and .898 CI_{95} (.758, 1.00) for post-secondary students. The effect size of the training at all school levels was medium.

Disabilities

Students participating in the studies were categorized into three disability groups—ASD, ID, and ASD + ID. The weighted Tau-*U* values of science education interventions were .910 CI_{95} (.740, 1.00) for the students with ASD, .810 CI_{95} (.729, .890) for the students with ID, and .816 CI_{95} (.717, .916) for the students with ASD + ID. The group that was least involved in the science education studies were students with ASD (9 AB pairs). The effect size interpretation of the science education studies was medium for all disability types. The effect size CI_{95} range for the students with ASD was smaller than the other disability categories (see Fig. 2).

Interventionists

Interventionist groups were placed into six categories—teachers, researchers, teachers and researchers, peers and researchers, facilitators and self-monitoring, and

Table 5 Effect Size Measurements of 18 Articles

Studies	**Number of AB Contrasts**	**Total Pairs**	**Number of Tau-U_{BC} Calculations** (If Tau-$U_A \geq .40$)	**Number of Tau-*U* Calculations**	**Weighted Tau-U (95% CI)**	**Effect Size**
Carnahan and Williamson (2013)	6	72	1	5	.947 (.563, 1.00)	Strong
Carnahan et al. (2016)	3	94	2	1	.925 (.504, 1.00)	Medium
Collins et al. (2011)	6	1483	2	4	.595 (.364, .825)	Small
Greene and Bethune (2019)	9	814	3	6	.516 (.319, .710)	Small
Heinrich et al. (2016)	2	60	0	2	.734 (.224, 1.00)	Medium
Hudson et al. (2014)	3	398	0	3	.604 (.320, .889)	Small
Karl et al. (2013)	4	202	0	4	.934 (.564, 1.00)	Strong
Knight et al. (2013)	3	324	0	3	.970 (.666, 1.00)	Strong
Knight et al. (2018a)	32	1446	3	29	.821 (.699, .944)	Medium
Knight et al. (2018b)	4	198	1	3	.913 (.590, 1.00)	Medium
McMahon et al. (2016)	24	732	9	15	.877 (.722, 1.00)	Medium
Miller and Taber-Doughty (2014)	3	45	1	2	.955 (.449, 1.00)	Strong
Miller et al. (2015)	6	116	0	6	.893 (.560, 1.00)	Medium
Riggs et al. (2013)	9	517	3	6	.693 (.457, .920)	Medium
Rivera et al. (2017)	3	140	0	3	1.00 (.620, 1.00)	Strong
Roberts et al. (2020)	3	380	0	3	.984 (.691, 1.00)	Strong
Smith et al. (2013a)	12	441	0	12	.898 (.699, 1.00)	Medium
Wood et al. (2020)	9	1022	1	8	.853 (.679, 1.00)	Medium
Total/ Aggregated Weighted Tau-*U*	141	8484	26	115	.820 (.762, .879)	Medium

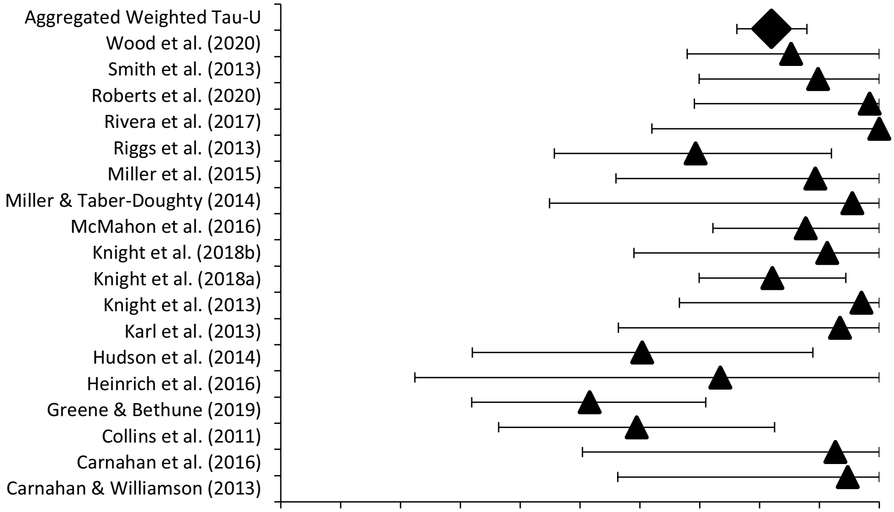

Fig. 3 Forest Plot for Individual Studies for Aggregated Weighted Tau-U and 95% Confidence Interval

online applications. The weighted Tau-*U* values of science education interventions were .808 CI_{95} (.723, .893) for teachers, .811 CI_{95} (.686, .936) for researchers, .853 CI_{95} (.679, 1.00) for teachers and researchers, .604 CI_{95} (.320, .889) for peers and researchers, .912 CI_{95} (.635, 1.00) for facilitators and self-monitoring, and .877 CI_{95} (.722, 1,00) for online applications. While the most employed operators were teachers (69 AB contrasts), researchers (27 AB contrasts) and the online applications (24 AB contrasts), the least employed operators were teachers and researchers (9 AB contrasts), facilitators and self-monitoring (9 AB contrasts), and peers and researchers (3 AB contrasts). Science education studies conducted with all interventionists except for peers and researchers (small effect) had a medium effect. The effect size CI_{95} range of teachers was smaller than other interventionists. On the other hand, the CI_{95} range of peers and researchers was larger than others.

Discussion

A descriptive, visual, and meta-analysis of the SCEDs teaching science education to students with IDD were analyzed in the current study. We aimed to answer the following research questions: (a) what skills were taught in science education studies, (b) what were the instructional approaches or adaptations used to teach science education to students with IDD, (c) what are the visual and effect size of SCEDs science education studies for students with IDD that met the CEC *QI*s (Cook et al., 2015):, and (d) what are the differences in these effect sizes, if any, depending on the participant and intervention characteristics.

Our systematic review yielded a total of 27 studies focused on improving science content knowledge and practices of students with IDD over the last 10 years. Generally speaking, the quality analysis of the studies using the *QI*s outlined by Cook

Table 6 Effect Size Measurements for Moderators

Moderator Variables	Number of AB Contrasts	Total Pairs	Number of Tau-U_{BC} Calculations (If Tau-$U_A \geq .40$)	Number of Tau-U Calculations	Weighted Tau-U (95% CI)	Effect Size
Participant Characteristics						
Grade Level						
Elementary (6–11)	56	3820	7	49	.764 (.599, .882)	Medium
Secondary (11–18)	57	3734	9	48	.838 (.739, .937)	Medium
Post-Secondary (18+)	28	930	10	18	.898 (.758, 1.00)	Medium
Disabilities						
ASD	9	166	3	6	.910 (.740, 1.00)	Medium
ID	72	4845	10	62	.810 (.729, .890)	Medium
ASD + ID	60	3473	13	47	.816 (.717, .916)	Medium
Intervention Characteristics						
Independent Variables						
Systematic Instruction	99	7425	13	86	.789 (.724, .854)	Medium
Self-Directed Learning	33	893	10	23	.900 (.770, 1.00)	Medium
Comprehension Based Learning	9	166	3	6	.938 (.647, 1.00)	Strong
Interventionists						
Teachers	69	4452	12	57	.808 (.723, .893)	Medium
Researchers	27	1719	3	24	.811 (.686, .936)	Medium
Teachers & Researchers	9	1022	1	8	.853 (.679, 1.00)	Medium
Peer(s) & Researchers	3	398	0	3	.604 (.320, .889)	Small
Facilitators and Self-Monitoring	9	161	1	8	.912 (.635, 1.00)	Medium
Online Applications	24	732	9	15	.877 (.722, 1.00)	Medium
Total/ Aggregated Weighted Tau-*U*	141	8484	26	115	.820 (.762, .879)	Medium

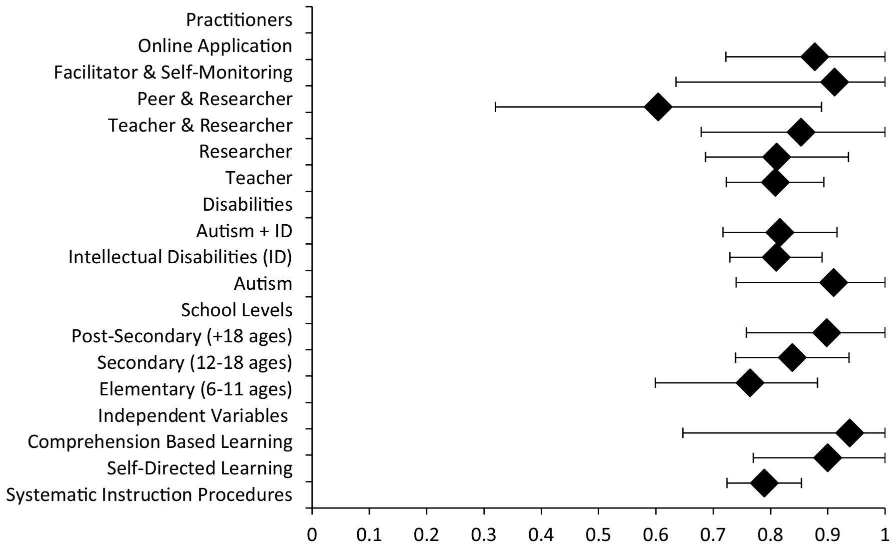

Fig. 4 Forest Plot for Moderating Variables for Aggregated Weighted Tau-U and CI_{95}

et al. (2015) aligns with previous quality review studies (i.e., Apanasionok et al., 2019; Knight et al., 2020) in that overall quality of science education studies for students with IDD appeared to be evidence-based practices. Of the 27 initial studies, 18 studies met all the CEC *QI*s; thus, further descriptive, visual, and meta-analysis were carried out for these 18 studies only.

Science vocabulary skills were taught across 6 of the 18 studies. Of the 6, Knight et al. (2018b) also taught comprehension and application of science vocabulary. Although our descriptive analysis aligned with the previous literature showing that the most frequently taught science knowledge is vocabulary (e.g., Apanasionok et al., 2019), our review also found that science *practices,* inquiry skills, and science text comprehension were taught across studies that met the CEC *QI*s (Cook et al., 2015). Future studies should continue to focus on more complex science knowledge and skills as opposed to only focusing on teaching basic vocabulary skills.

Of the 18 studies included in this review, 13 used explicit and systematic instructional approaches, and the outcomes were positive in all studies on teaching science vocabulary. Given the issue stated above regarding the need to move beyond science vocabulary, future studies should focus on the most efficacious instructional approaches that go beyond teaching basic skills. Explicit and systematic instruction can certainly meet this need. However, there is a clear need in the research literature to determine if popular approaches to science (e.g., problem-based learning) can or should be integrated with explicit and systematic approaches to teaching more advanced science content for students with IDD in the general education classroom. For example, it is important to determine whether or not after critical prerequisite skills are taught through explicit and systematic instructional methods, there is value in adding these other approaches afterward to aid with and/or to assess generalization to new and novel science content or conditions.

In addition to the method of instruction, Osborne (2007) indicated that a critical issue in science education in the twenty-first century is the type of assessment used to measure science knowledge. Given that SCEDs require frequent repeated measures (Martella et al., 2019), it is not surprising that researchers used quizzes or task analyses where participants were expected to choose a correct response among two to three incorrect responses, verbally respond to questions, or place a picture of correct answers in an array. Measures of comprehension or functional application of science concepts often were not measured in the studies that targeted science vocabulary. Further, five studies that targeted science vocabulary provided assessments on the same questions taught but only in a different order. Although Spooner et al. (2011) stated over a decade ago that future science vocabulary studies would not only focus on the definition of the words but also build hands-on activities to apply such skills, it appears that there is still needed improvement in this area. Thus, future researchers should include ongoing measurements to assess comprehension (e.g., Knight et al., 2018b) and functional application (e.g., Collins et al., 2011) of science concepts to determine the level of science knowledge due to instructional programs. Additionally, researchers ought to consider integrating standardized assessments into their investigations (e.g., Kamps et al., 2016) when using explicit, systematic, and scripted (ESS) programs since such an approach can strengthen the determination of the effectiveness of an intervention (Sulu et al., 2021).

Although there is research support for the acquisition and fluency of science knowledge, there is a paucity of research on the long-term maintenance and generalization. The successful acquisition of behavior does not guarantee that the same behavior will be maintained across time nor be generalized to different situations (Cooper et al., 2020; Phillips & Vollmer, 2012; Schreibman, 2005). Our findings show that the demonstration of maintenance and generalization of science knowledge was limited. Long-term maintenance data were not collected after the completion of the intervention in any of the reviewed studies. The greatest duration of maintenance was 6 weeks across all the studies.

Additionally, the generalized effects of science education instruction were not systematically assessed across many of the studies. Students often learn a new skill under a certain set of conditions that may include specific persons, materials, and in specific environments where they learn these skills (Baer et al., 1968); however, there is a need for further and specific planning for generalization of skills beyond instructional settings (Marchand-Martella & Martella, 2002; Neely et al., 2018) to say a behavioral change has taken place (Baer et al., 1968). Thus, explicit programming across novel settings, individuals, and behaviors is needed for generalization to occur (Marchand-Martella et al., 2013; Baer et al., 1968; Handleman, 1979; Stokes & Baer, 1977). Of the 18 studies reviewed, 11 of them assessed for generalization. Excluding Knight et al. (2013), none of the studies used an additional teaching strategy. Furthermore, generalization results yielded negative results in 2 studies. Therefore, future studies should program for and assess the maintenance and generalization of learned science content and practices for students with IDD.

Another area that needs to be improved is that science studies were predominantly implemented in segregated settings (e.g., resource rooms). Evidence-based practices are insufficient unless the delivery of these practices takes place in general education environments (Agran et al., 2020). Additionally, one of the requirements in the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) is the utilization of specifically designed instructions to support students with disabilities (Yell et al., 2020) so that all students can make progress in the general education curriculum within the least restrictive environments. Therefore, there is a need for future studies to determine the educational impact of instructional programs on the acquisition of science content in general education classroom settings.

Visual and Effect Size Analysis

The visual analyses of the 18 studies included therapeutic trends for all the studies included in this analysis and these improvements were at varying ratios. Although visual analysis is a predominant method in analyzing of SCEDs, it is not without limitations. For example, whether the amount of level change or immediacy of effect is significant cannot be evaluated via visual analysis. A trend can be in the targeted direction; however, the significance of this direction is evaluated based on the practitioner's interpretation. Therefore, visual analysis is criticized for allowing such subjective evaluations (Aydin & Tanious, 2022; Brossart et al., 2014; Kratochwill et al., 2014; Harrington & Velicer, 2015; Ninci et al., 2015).

On the other hand, studies can be analyzed and interpreted objectively with statistical analyses such as effect sizes. In the present study, we sought to identify the overall effect of studies meeting the CEC *QI* standards using single-case meta-analysis in addition to visual analysis. The data from 141 AB graphs across 18 studies were analyzed, and results revealed that the interventions generated a medium overall effect (Tau-U = .82, CI_{95} [.76, .87]). These data provide preliminary evidence that different science education interventions are effective in improving the science content knowledge and practices of students with IDD. Our results align with the previous meta-analysis of the overall effect size of science education studies. Therrien et al. (2011) indicated that science education studies have moderate effect sizes for students with learning disabilities. Similarly, Therrien et al. (2014) found the science instruction studies to have a small to moderate impact on students with emotional and behavioral disorders. Although group experimental studies and/or SCEDs were included in these reviews, and different effect size methods (e.g., g statistic, PND, PAND) were used, the present study found similar results.

There were no significant differences across the moderator variables of grade levels (i.e., elementary, secondary, and post-secondary) and disability categories (i.e., ASD, ID, and ASD + ID). For these moderator variables, the effect sizes were medium. On the other hand, comprehension-based learning had a strong effect with a large CI_{95} range, whereas explicit and systematic instruction and self-directed learning had a medium effect with a small CI_{95} range. Thus, comprehension-based learning may be more effective than explicit and systematic, and self-directed learning methods; however, there is a need for further studies to validate these findings, given that only two of the 18 studies used comprehension-based learning. For interventionists, the effect sizes for all except one (i.e., peer and researcher) had a medium effect demonstrating that the result did not change in the presence of different people implementing science instruction.

Overall, the findings showed that the greater the level of change, the immediacy of effect, and the consistency rates were, the greater the effect size was. In addition, the lower the overlap percentage was, the higher the effect size was. Although there is a consistency between visual analysis and effect size indicators, we argue that visual analysis cannot be used alone in evaluating the effectiveness of science education studies. Because the effect size can produce a single value by combining individual effect sizes of studies and allowing the overall result of science education studies, this is not possible in visual analysis. In addition, visual analysis may involve subjective assessments among coders. This finding supports the literature advocating that visual analysis can be complemented by statistical analyses (Aydin & Tanious, 2022; Harrington & Velicer, 2015; Kratochwill et al., 2014; Manolov & Vannest, 2019; Vannest & Ninci, 2015). It should also be noted that the sample of moderator variables (the number of AB contrasts) was small, especially when the moderator variables (i.e., peer and researcher, comprehension-based learning, ASD) had an effect size with a large CI_{95} range. Therefore, there is a need for further studies to determine more precise and credible results among these moderators.

Limitations

Although this meta-analysis found that the current research on teaching science content and practices to students with IDD meets *QIs* and found these interventions are effective in overall effect size, it was not without limitations. First, our search included studies up to August 2021, so there might have been other studies published after our initial search. Second, one inclusion criterion was a diagnosis of IDD for at least one of the participants within a study; therefore, studies may have been excluded due to an absence of participant diagnosis. Third, we only included SCEDs published in peer-review journals, given that SCEDs are predominantly used in science education studies to teach students with IDD (e.g., Knight et al., 2020). Fourth, we used CEC *QI*s to evaluate the overall quality of studies before conducting visual and meta-analyses, given that CEC *QI*s were developed based on the limitations of previous quality assessments and commonly used in the literature (e.g., David et al., 2022). However, we acknowledge that different criteria published by different authors (e.g., Horner et al., 2005) may yield different results. Fifth, we used a nonoverlap-based method (i.e., Tau-*U*) to calculate the effect sizes of the studies and moderators. Regarding nonoverlap-based methods, there are some unresolved issues, such as invalid results in the case of outliers and the inability to determine the magnitude of socially meaningful behavior change (Aydin & Tanious, 2022). Therefore, future meta-analyses should consider utilizing recent indexes such as performance criteria-based effect size (PCES; Aydin & Tanious, 2022). Finally, gray literature (e.g., book chapters, dissertation) was excluded because the literature review was conducted in peer-review journals only. Excluding gray literature may have caused publication bias and unpublished studies would have small effect sizes (Polanin et al., 2016). Thus, future studies should consider including gray literature and analyze the gray literature as a variable of analysis (Lory et al., 2020).

Recommendations for Future Research

Science education plays a significant impact on an individual's independence within society and opens more opportunities, including jobs in science-related practices (Ehsan et al., 2018; Rizzo & Taylor, 2016). Providing students with IDD with effective instructions and support would increase their critical thinking, problem-solving, and employability (Ehsan et al., 2018; Rizzo & Taylor, 2016). Although the emphasis on science education to teach students with IDD has increased in the last decade, the research in teaching science content and practices are still limited compared to the other content areas such as math and literacy. The small number of studies included in this analysis revealed that the scope of the targeted skills is inadequate to support the aforementioned skills (e.g., employability) of students with IDD. Upon considering the changing trends in teaching science education for the twenty-first century (Forbes et al., 2020; Osborne, 2007), future studies should focus on going beyond basic skills, ensuring comprehension and application of these skills.

In addition to the limitations in the scope of the literature, the assessment of student outcomes was largely measured based on experimenter-designed assessments only. Given that the majority of studies were carried out in special education settings, it is unclear if the similar outcomes would have been measured in typical instructional settings or general education classrooms (Rizzo & Taylor, 2016; Taylor et al., 2020). Therefore, there is a need for future studies where experimenter-designed probes are combined with standardized tests to measure students' outcomes. Further, such measurements ought to be addressed in the typical instructional settings or general education classrooms because all students should have access to learning all scientific content and practices in general education classrooms. Furthermore, future studies should provide the level of participants' background knowledge of science concepts prior to interventions and program for and assess the maintenance and generalization of learned science content in, for example, real-life circumstances (Spooner et al., 2011).

One other area that needs to be improved is the diversity of instructional practices in teaching students with IDD. Although inquiry-based instruction is the preferred method of science instruction among science educators to teach students without disabilities (Taylor et al., 2020), individuals with IDD were predominantly taught with explicit and systematic instructions. Future studies should consider the efficacy of using inquiry-based learning or problem-based learning approaches along with explicit and systematic approaches. Based on the limited number of studies included in this analysis, there is also a need to conduct more research on science content instruction to determine more the effects of different instructional approaches alone or in combination for students with IDD.

Recommendations for Practitioners

The role of teacher training is significant to instructional practices in general education classrooms (Sulu et al., 2021; Rizzo & Taylor, 2016). If our ultimate goal is to integrate students with IDD into the general educational setting, practitioners must begin to move beyond mainly teaching science vocabulary and memorization of science facts. Although science vocabulary and facts are critical for laying the foundation of science literacy, practitioners must begin to move into more complex areas of science education that involve science content knowledge and problem-solving. Doing so would be the next step in the evolution of teaching complex science content and problem-solving to students with IDD.

A critical issue in teaching students with IDD more complex skills is the instructional methods used. Unfortunately, a dichotomy of instructional methods exists, putting explicit and systematic instruction against inquiry-based methods. The question is not which to use but when to use each one. Practitioners should begin to integrate these instructional approaches based on the best evidence in science instruction. For example, basic skills such as vocabulary and facts and other foundational knowledge required in science can be taught using an explicit and systematic instructional approach. Once these foundational skills have been firmed, inquiry-based methods may begin to be utilized for more complex science concepts and problem-solving.

Additionally, it is important for practitioners to determine if their instruction is effective in teaching science content. Frequent measures should be used to track student progress toward science content goals and objectives. Doing so allows practitioners to determine if students are making progress and, if not, make changes in a timely manner to enhance the acquisition of science skills and knowledge.

Finally, if the goal for students with IDD is to integrate them into general education settings, generalization programming should take place when instruction is occurring. Instructional methods used in general education environments may need to be phased in during science instruction (i.e., cooperative learning, problem-based learning) so students with IDD are not without experience with such approaches. Additionally, some form of assessment should be used to determine if students with IDD are successful in future learning environments where science education is taking place, and the use of remedial instructional approaches are developed and implemented if/when students with IDD begin to struggle with science concepts.

Funding This study was not supported with any funding.

Data Availability All original documents can be provided by the corresponding author.

Compliance with Ethical Standards

Conflict of Interest The authors have no conflicts of interest to declare.

Ethical Approval This article does not contain any studies with human participants or animal performed by any of the authors.

Informed Consent This article does not contain any studies with human participants or animal performed by any of the authors.

References

*Articles Included in Meta-analysis

Agran, M., Jackson, L., Kurth, J. A., Diane, R., Burnette, K., Jameson, M., Zagona, A., Fitzpatrick, H., & Wehmeyer, M. (2020). Why aren't students with severe disabilities being placed in general education classrooms: Examining the relations among classroom placement, learner outcomes, and other factors. *Research and Practice for Persons with Severe Disabilities, 45*, 4–13. https://doi.org/10.1177/2F1540796919878134

Apanasionok, M. M., Hastings, R. P., Grindle, C. F., Watkins, R. C., & Paris, A. (2019). Teaching science skills and knowledge to students with developmental disabilities: A systematic review. *Journal of Research Science Teaching, 56*, 847–880. https://doi.org/10.1002/tea.21531

Aydin, O., & Tanious, R. (2022). Performance criteria-based effect size (PCES) measurement of single-case experimental designs: A real-world data study. *Journal of Applied Behavior Analysis, 55*(3), 891–918. https://doi.org/10.1002/jaba.928

Aydin, O., & Yassikaya, M. Y. (2022). Validity and reliability analysis of the PlotDigitizer software program for data extraction from single-case graphs. *Perspectives on Behavior Science, 45*(1), 239–257. https://doi.org/10.1007/s40614-021-00284-0

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–97. https://doi.org/10.1901/jaba.1968.1-91

Barnett, J. H., Trillo, R., & More, C. (2018). Visual supports to promote science discourse for middle and high school students with autism spectrum disorders. *Intervention in School and Clinic, 53*, 292–299. https://doi.org/10.1177/1053451217736865

Barthlow, M. J., & Watson, S. B. (2014). The effectiveness of process-oriented guided inquiry learning to reduce alternative conceptions in secondary chemistry. *School Science and Mathematics, 114*(5), 246–255. https://doi.org/10.1111/ssm.12076

Brigham, F. J., Scruggs, T. E., & Mastropieri, M. A. (2011). Science education and students with learning disabilities. *Learning Disabilities Research & Practice, 26*(4), 223–232. https://doi.org/10.1111/j.1540-5826.2011.00343.x

Brossart, D. F., Laird, V. C., & Armstrong, T. W. (2018). Interpreting Kendall's Tau and Tau-U for single-case experimental designs. *Cogent Psychology, 5*(1), 1–26. https://doi.org/10.1080/23311908.2018.1518687

Brossart, D. F., Vannest, K. J., Davis, J. L., & Patience, M. A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation, 24*(3–4), 464–491. https://doi.org/10.1080/09602011.2013.868361

Browder, D. M., & Spooner, F. (2011). *Teaching students with moderate and severe disabilities.* Guilford Publications.

*Carnahan, C. R., & Williamson, P. S. (2013). Does compare-contrast text structure help students with autism spectrum disorder comprehend science text? *Council for Exceptional Children, 79*(3), 347–363. https://doi.org/10.1177/2F001440291307900302.

*Carnahan, C. R., Williamson, P., Birri, N., Swoboda, C., & Snyder, K. K. (2016). Increasing comprehension of expository science text for students with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, 31(3), 208–220. https://doi.org/10.1177/2F1088357615610539.

Cawley, J. F (1994). Science for Students with Disabilities. *Remedial and Special Education*, *15*(2), 67–71. https://doi.org/10.1177/2F074193259401500202

*Collins, B., Hager, K., & Galloway, C. (2011). Addition of functional content during core content instruction with students with moderate disabilities. *Education and Training in Autism and Developmental Disabilities, 46*, 22–39. http://www.jstor.org/stable/23880028

*Collins, B. C., Terrell, M., & Test, D. W. (2017). Using a simultaneous prompting procedure to embed core content when teaching a potential employment skill. *Career Development and Transition for Exceptional Individuals*, *40*, 36–44. https://doi.org/10.1177/2F2165143416680347

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36*(4), 220–234. https://doi.org/10.1177/0741932514557271

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis.* Russell Sage Foundation.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis (3rd ed.).* Pearson.

David, M., Mason, R. A., Davis, J. L., Gregori, E., Lei, Q., Lory, C., & Wang, D. (2022). School-based interventions targeting challenging behavior of adolescents with developmental disabilities: A meta-analysis. *Journal of Developmental and Physical Disabilities*, 1–30. https://doi.org/10.1007/s10882-022-09841-0

Dowdy, A., Peltier, C., Tincani, M., Schneider, W. J., Hantula, D. A., & Travers, J. C. (2021). Meta-analyses and effect sizes in applied behavior analysis: A review and discussion. *Journal of Applied Behavior Analysis, 54*(4), 1317–1340. https://doi.org/10.1002/jaba.862

Ehsan, H., Rispoli, M., Lory, C., & Gregori, E. (2018). A systematic review of STEM instruction with students with autism spectrum disorders. *Review Journal of Autism and Developmental Disorders, 5*(4), 327–348. https://doi.org/10.1007/s40489-018-0142-8

*Fetko, E., Collins, B., Hager, K., & Spriggs, A. (2013). Embedding science facts in leisure skill instruction conducted by peer tutors. *Education and Training in Autism and Developmental Disabilities, 48*, 400–411. http://www.jstor.org/stable/23880996

Forbes, C. T., Neumann, K., & Schiepe-Tiska, A. (2020). Patterns of inquiry-based science instruction and student science achievement in PISA 2015. *International Journal of Science Education, 42*(5), 783–806. https://doi.org/10.1080/09500693.2020.1730017

*Greene, A., & Bethune, K. S. (2019). The effects of systematic instruction in a group format to teach science to students with autism and intellectual disability. *Journal of Behavioral Education, 30*, 62–79. https://doi.org/10.1007/s10864-019-09353-6

Handleman, H. S. (1979). Generalization by autistic-type children of verbal responses across settings. *Journal of Applied Behavior Analysis, 12*, 273–284. https://doi.org/10.1901/jaba.1979.12-273

Harrington, M., & Velicer, W. F. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research, 50*(2), 162–183. https://doi.org/10.1080/00273171.2014.973989

*Heinrich, S., Collins, B., Knight, V., & Spriggs, A. (2016). Embedded simultaneous prompting procedure to teach stem content to high school students with moderate disabilities in an inclusive setting. *Education and Training in Autism and Developmental Disabilities, 51*, 41–54. https://www.jstor.org/stable/26420363

Horner, R. H., Carr, E. G., Halle, J., McGee, G. G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. https://doi.org/10.1177/001440290507100203

*Hudson, M., Browder, D., & Jimenez, B. (2014). Effects of a peer-delivered system of least prompts intervention and adapted science read-alouds on listening comprehension for participants with moderate intellectual disability. *Education and Training in Autism and Developmental Disabilities, 49*, 60–77. http://www.jstor.org/stable/23880655

Individuals With Disabilities Education Improvement Act, 20 U.S.C. § 1400 (2004). Retrieved January 21, 2023, from https://sites.ed.gov/idea/

*Jimenez, B., A, Browder, D. M., Spooner, F., & Dibiase, W. (2012). Inclusive inquiry science using peer-mediated embedded instruction for students with moderate intellectual disability. *Exceptional Children*, *78*, 301–317. https://doi.org/10.1177/2F001440291207800303

*Jimenez, B. A., Lo, Y., & Saunders, A., F. (2014). The additive effects of scripted lessons plus guided notes on science quiz scores of students with intellectual disability and autism. *The Journal of Special Education*, *47*, 231–244. https://doi.org/10.1177/2F0022466912437937

Kaldenberg, E. R., Watt, S. J., & Therrien, W. J. (2015). Reading instruction in science for students with learning disabilities: A meta-analysis. *Learning Disability Quarterly*, *38*(3), 160–173. https://doi.org/10.1177/2F0731948714550204

Kamps, D., Heitzman-Powell, L., Rosenberg, N., Mason, R., Schwartz, I., & Romine, R. (2016). Effects of Reading Mastery as a small group intervention for young children with ASD. *Journal of Developmental and Physical Disabilities, 28*(5), 703–722. https://doi.org/10.1007/s10882-016-9503-3

*Karl, J., Collins, B., Hager, K., & Ault, M. (2013). Teaching core content embedded in a functional activity to students with moderate intellectual disability using a simultaneous prompting procedure. *Education and Training in Autism and Developmental Disabilities, 48*, 363–378. http://www.jstor.org/stable/23880993

Kennedy, C. H. (2005). *Single-case designs for educational research*. Allyn & Bacon.

*Knight, V. F., Collins, B., Spriggs, A. D., Sartini, E., & MacDonald, M. J. (2018a). Scripted and unscripted science lessons for children with autism and intellectual disability. *Journal of Autism Developmental Disorders, 48*, 2542–2557. https://doi.org/10.1007/s10803-018-3514-0

*Knight, V. F., Creech-Galloway, C. E., Karl, J. M., & Collins, B. C. (2018b). Evaluating supported eText to teach science to high school students with moderate intellectual disability. *Focus on Autism and Other Developmental Disabilities*, *33*, 227–236. https://doi.org/10.1177/2F1088357617696273

*Knight, V. F., Smith, B. R., Spooner, F., & Browder, D. (2012). Using explicit instruction to teach science descriptors to students with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 42*, 378–389. https://doi.org/10.1007/s10803-011-1258-1

*Knight, V. F., Spooner, F., Browder, D. M., Smith, B. R., & Wood, C. L. (2013). Using systematic instruction and graphic organizers to teach science concepts to students with autism spectrum disorders and intellectual disability. *Focus on Autism and Other Developmental Disabilities, 28*, 115–126. https://doi.org/10.1177/1088357612475301

Knight, V. F., Wood, L., McKissick, B. R., & Kuntz, E. M. (2020). Teaching science content and practices to students with intellectual disability and autism. *Remedial and Special Education*, *41*, 327–340. https://doi.org/10.1177/2F0741932519843998

*Knight, V. F., Wood, C. L., Spooner, F., Browder, D. M., & O'Brien, C. P. (2015). An exploratory study using science eTexts with students with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities, 30*, 86–99. https://doi.org/10.1177/2F1088357614559214

Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological issues. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advantages,* (pp. 91–126). American Psychological Association. https://doi.org/10.1037/14376-004

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3–4), 445–463. https://doi.org/10.1080/09602011.2013.815636

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed.). Routledge

Lory, C., Mason, R. A., Davis, J. L., Wang, D., Kim, S. Y., Gregori, E., & David, M. (2020). A meta-analysis of challenging behavior interventions for students with developmental disabilities in inclusive school settings. *Journal of Autism and Developmental Disorders, 50*(4), 1221–1237. https://doi.org/10.1007/s10803-019-04329-x

Maggin, D. M., Cook, B. G., & Cook, L. (2018). Using single-case research designs to examine the effects of interventions in special education. *Learning Disabilities Research & Practice, 33*(4), 182–191. https://doi.org/10.1111/ldrp.12184

Manolov, R., & Vannest, K. J. (2019). A visual aid and objective rule encompassing the data features of visual analysis. *Behavior Modification*, 0145445519854323. https://doi.org/10.1177/0145445519854323.

Marchand-Martella, N., & Martella, R. (2002). An overview and research summary of peer-delivered corrective reading instruction. *The Behavior Analyst Today, 3*(2), 213–220. https://doi.org/10.1037/h0099966

Marchand-Martella, N. E., Martella, R. C., Modderman, S. L., Petersen, H. M., & Pan, S. (2013). Key areas of effective adolescent literacy programs. *Education and Treatment of Children, 36*(1), 161–184. https://doi.org/10.1353/etc.2013.0005

Martella, R. C., Nelson, J. R., Marchand-Martella, N. E., & Sinclair, T. E. (2019). The quality of single-case evaluation studies of curricular programs for students with disabilities. *International Journal of Special Education, 34*(1), 1–16. Retrieved January 21, 2023, from https://eric.ed.gov/?id=EJ1237129

Martella, R. C., Nelson, J. R., Morgan, R. L., & Marchand-Martella, N. E. (2013). *Understanding and interpreting educational research.* Guilford Press.

Mason, L. H., & Hedin, L. R. (2011). Reading science text: Challenges for students with learning disabilities and considerations for teachers. *Learning Disabilities Research & Practice, 26*, 214–222. https://doi.org/10.1111/j.1540-5826.2011.00342.x

Mastropieri, M. A., & Scruggs, T. E. (1992). Science for students with disabilities. *Review of Educational Research*, *62*(4), 377–411. https://doi.org/10.3102/2F00346543062004377

Mastropieri, M. A., Scruggs, T. E., Norland, J. J., Berkeley, S., McDuffie, K., Tornquist, E. H., & Connors, N. (2006). Differentiated curriculum enhancement in inclusive middle school science: Effects on classroom and high-stakes tests. *The Journal of Special Education, 40*(3), 130–137. https://doi.org/10.1177/00224669060400030101

*McKissick, B.R., Davis, L.L., Spooner, F., Fisher, L.B., & Graves, C. (2018). Using computer-assisted instruction to teach science vocabulary to students with autism spectrum disorder and intellectual disability. *Rural Special Education Quarterly*, *37*(4), 207–218. https://doi.org/10.1177/2F8756870518784270

*McMahon, D. D., Cihak, D. F., Wright, R. E., & Bell, S. M. (2016). Augmented reality for teaching science vocabulary to postsecondary education students with intellectual disabilities and autism. *Journal of Research on Technology in Education, 48*(1), 1–19. https://doi.org/10.1080/15391523.2015.1103149

Meij, H., Meij, J., & Harmsen, R. (2015). Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educational Technology Research and Development, 63*(3), 381–403. https://doi.org/10.1007/s11423-015-9378-5

*Miller, B., & Taber-Doughty, T. (2014). Self-monitoring checklists for inquiry problem-solving: Functional problem-solving methods for students with intellectual disability. *Education and Training in Autism and Developmental Disabilities, 49*(4), 555–567. https://www.jstor.org/stable/24582351.

*Miller, B., Doughty, T., & Krockover, G. (2015). Using science inquiry methods to promote self-determination and problem-solving skills for students with moderate intellectual disability. *Education and Training in Autism and Developmental Disabilities*, *50*(3), 356–368. http://www.jstor.org/stable/24827516.

National Center for Education Statistics. (2015). *National assessment of educational progress 4th grade science assessment*. Institute of Education Sciences.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (Committee on a Conceptual Framework for New K-12 Science Education Standards, Board on Science Education, Division of Behavioral and Social Sciences and Education). The National Academies Press.

Neely, L. C., Garcia, E., Bankston, B., & Green, A. (2018). Generalization and maintenance of functional communication training for individuals with developmental disabilities: A systematic and quality review. *Research in Developmental Disabilities, 79*, 116–129. https://doi.org/10.1016/j.ridd.2018.02.002

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Washington, DC: The National Academic Press.

Ninci, J., Vannest, K. J., Willson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification, 39*(4), 510–541. https://doi.org/10.1177/0145445515581327

Osborne, J. (2007). Science Education for the Twenty First Century. *Eurasia Journal of Mathematics, Science & Technology Education, 3*(3), 173–184. https://doi.org/10.12973/ejmste/75396

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education, 25*, 177–196. https://doi.org/10.1007/s10972-014-9384-1

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-*U*. *Behavior Therapy, 42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Phillips, C. L., & Vollmer, T. R. (2012). Generalized instruction following with pictorial prompts. *Journal of Applied Behavior Analysis, 45*, 37–54. https://doi.org/10.1901/jaba.2012.45-37

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, *86*(1), 207–236. https://doi.org/10.3102/2F0034654315582067

*Riggs, L, Collins, B. C., Kleinert, H., & Knight, V. F. (2013). Teaching principles of heredity to high school students with moderate and severe disabilities. *Research and Practice for Persons with Severe Disabilities*, *38*, 30–43. https://doi.org/10.2511/2F027494813807046971

*Rivera, C. J., Hudson, M. E., Weiss, S. L., & Zambone, A. (2017). Using a multicomponent multimedia shared story intervention with an iPad to teach content picture vocabulary to students with developmental disabilities. *Education & Treatment of Children, 40*, 327–352. https://doi.org/10.1353/etc.2017.0014

Rizzo, K. L., & Taylor, J. C. (2016). Effects of inquiry-based learning instruction on science achievement for students with disabilities: An analysis of the literature. *Journal of Science Education for Students with Disabilities, 19*, 1–16.

*Roberts, C. A., Tandy, J., Kim, S. Y., & Meyer, N. (2020). A multi-component literacy intervention with science expository text for students with moderate intellectual disability. *Education and Training in Autism and Developmental Disabilities, 55*, 382–397. https://www.proquest.com/scholarly-journals/multi-component-literacy-interventionwith/docview/2463177375/se-2?accountid=13360

Schreibman, L. (2005). *The science and fiction of autism*. Harvard University Press.

Scruggs, T. E., & Mastropieri, M. A. (2007). Science learning in special education: The case for constructed versus instructed learning. *Exceptionality, 15*(2), 57–74. https://doi.org/10.1080/09362830701294144

Scruggs, T. E., Mastropieri, M. A., & Boon, R. (1998). Science education for students with disabilities: A review of recent research. *Studies in Science Education, 32*(1), 21–44. https://doi.org/10.1080/03057269808560126

*Smith, B. R., Spooner, F., Jimenez, B. A., & Browder, D. (2013a). Using an Early Science curriculum to teach science vocabulary and concepts to students with severe developmental disabilities. *Education & Treatment of Children, 36*, 1–31. https://www.jstor.org/stable/42900600

*Smith, B. R., Spooner, F., & Wood, C. L. (2013b). Using embedded computer-assisted explicit instruction to teach science to students with autism spectrum disorder. *Research in Autism Spectrum Disorder, 7*, 433–443. https://doi.org/10.1016/j.rasd.2012.10.010

Spooner, F., Knight, V., Browder, D., Jimenez, B., & DiBiase, W. (2011). Evaluating evidence-based practice in teaching science content to students with severe developmental disabilities. *Research & Practice for Persons with Severe Disabilities, 36*, 62–75. https://doi.org/10.2511/2Frpsd.36.1-2.62

Spooner, F., McKissick, B. R., & Knight, V. F. (2017). Establishing the state of affairs for evidence-based practices in students with severe disabilities. *Research and Practice for Persons with Severe Disabilities*, *42*, 8–18. https://doi.org/10.1177/2F1540796916684896

*Starling, A., Peirce, L., Lo, Y., & Christopher, J. (2015, Winter). Improving science scores of middle school students with learning disabilities through engineering problem solving activities. *Journal of the American Academy of Special Education Professionals. 98*–113. https://eric.ed.gov/?id=EJ1134214

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349–367. https://doi.org/10.1901/jaba.1977.10-349

Sulu, M. D., Martella, R. C., Grimmet, K., Borosh, A. M., & Erden, E. (2022). Investigating the effects of self-monitoring interventions with students with disabilities on the maintenance and generalization of on-task behavior: A systematic literature review. *Review Journal of Autism and Developmental Disorders,* 1–19. https://doi.org/10.1007/s40489-022-00304-y

Sulu, M. D., Martella, R. C., Toper, O., Marchand-Martella, N. E., & Kiyak, U. E. (2021). Explicit and systematic scripted instructional programs for students with autism spectrum disorder: An updated and extended review. *Review Journal of Autism and Developmental Disorders,* 1–17. https://doi.org/10.1007/s40489-021-00284-5

Taylor, J. C., Rizzo, K. L., Hwang, J., & Hill, D. (2020). A review of research on science instruction for students with autism spectrum disorder. *School Science and Mathematics, 120*(2), 116–125. https://doi.org/10.1111/ssm.12388

Therrien, W. J., Taylor, J. C., Hosp, J. L., Kaldenberg, E. R., & Gorsh, J. (2011). Science instruction for students with learning disabilities: A meta-analysis. *Learning Disabilities Research & Practice, 26*, 188–203. https://doi.org/10.1111/j.1540-5826.2011.00340.x

Therrien, W. J., Taylor, J. C., Watt, S., Kaldenberg, E. R. (2014). Science instruction for students with emotional and behavioral disorders. *Remedial and Special Education*, *35*(1), 15–27. https://doi.org/10.1177/2F0741932513503557

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development, 93*(4), 403–411. https://doi.org/10.1002/jcad.12038

Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). *Single Case Research: web based calculators for SCR analysis.* (Version 2.0) [Web-based application]. College Station, TX: Texas A&M University. Available from http://singlecaseresearch.org

Wei, X., Christiano, E. R., Yu, J. W., Wagner, M., & Spiker, D. (2015). Reading and math achievement profiles and longitudinal growth trajectories of children with an autism spectrum disorder. *Autism, 19*, 200–210. https://doi.org/10.1177/1362361313516549

What Works Clearinghouse. (2022). *Procedures Handbook (Version 5)*. Washington, DC: Institute of Educational Sciences, U. S. Department of Education.

*Wood, L, Browder, D. M, & Spooner, F. (2020). Teaching listening comprehension of science e-texts for students with moderate intellectual disability. *Journal of Special Education Technology*, *35*, 272–285. https://doi.org/10.1177/2F0162643419882421

Yell, M. L., Collins, J., Kumpiene, G., & Bateman, D. F. (2020). The individualized education program: Procedural and substantive requirements. *Teaching Exceptional Children, 52*, 304–318. https://doi.org/10.1177/2F0040059920906592