Invited Lecture

The Effects of Interventions with Mathematics Manipulatives on Generalization and Maintenance for Children with Autism Spectrum Disorder: A Meta-Analysis of Single-Case Studies

Di Liu[1], Yan Feng[1], Ziyi Chen[1], Rui Kang[2], and Yifan Zuo[1]

ABSTRACT There have been reviews or meta-analysis showing that using manipulatives is an effective intervention for learning mathematics for students with disabilities, including autism spectrum disorders (ASD), without concentrating on the effects on generalization and maintenance. We conducted a meta-analysis to evaluate the effect of manipulatives on generalizing and/or maintaining mathematical skills for individuals with ASD and whether the effect varies with different participant characteristics, study design, intervention characteristics and mathematical content, focusing on the single-case studies. After application of the What Works Clearinghouse design standards, a total of 11 studies were included in the review: three studies collected data points during generalization phases, five studies collected data points during maintenance phases, the other three studies collected both generalization and maintenance data. Aggregate Tau-U and non-overlap of all pairs effect sizes (NAP) were calculated for each study and conducted moderator analyses. Overall, effect size scores ranged from small to significant effects across all comparisons. On average, most comparisons from the baseline to generalization and maintenance produced medium to large effects. Whereas, minor effects were found in most of the intervention of generalization and maintenance comparisons. Further moderator analysis regard to generalization and maintenance revealed that out of seven variables analyzed, only manipulatives types served as a moderator for maintenance. The findings suggest that manipulatives interventions were likely to result in mixed effects on mathematical skill generalization and maintenance within children with ASD, especially virtual manipulatives. Limitations and implications for future research and practice are discussed.

Keywords: Manipulatives; Mathematics; Autism spectrum disorder; Generalization and maintenance; Single-case research; Meta-analysis.

1. Introduction

Manipulatives, one instructional approach, are widely used in mathematics classes (Carbonneau et al., 2013), defined as objects designed to represent explicitly and

[1] College of Teacher Education, East China Normal University, Shanghai, China
[2] Department of Teacher Education, Georgia College & State University, GA, USA

concretely mathematical ideas that are abstract (Moyer, 2001). They have both visual and tactile appeal and can be manipulated by learners through hands-on experience which supports students' conceptual understanding of mathematical content (Moyer, 2001). The National Council of Teachers of Mathematics (NCTM, 2000) has recommended to use manipulatives for students to develop mathematical understanding. Furthermore, systematic reviews on the use of manipulatives to improve mathematical outcomes for students indicated that using manipulatives was effective at improving mathematical outcomes (Carbonneau et al., 2013; Sowell, 1989).

Manipulatives come in two types: the physically represented form of concrete manipulatives, and the computer-generated form of virtual manipulatives (Bouck and Flanagan, 2010; Moyer et al., 2002). Concrete manipulatives are physical objects that are used to engage students in hands-on learning of mathematics to introduce math concepts. Examples of concrete manipulatives include Base 10 Blocks, Pattern Blocks, algebra tiles and fraction pieces. They allow students have numerous materials to manipulate and opportunity to sort, classify, weigh, stack and explore. Therefore, when students use concrete manipulatives to explore and master concepts, they were more engaged and motivated (Moyer, 2001). However, researchers pointed that there are two challenges associated with concrete manipulatives. One is the case that dealing with multiple physical pieces may distract students' thought process, and the other is multiple pieces may increase cognitive load, leading to a lack of mathematical concepts (Suh and Moyer, 2008). The increased presence of technology in today's classrooms supports an exploration of virtual manipulatives (Bouck et al., 2018) which may represent an appropriate substitute for concrete manipulatives (Bassette et al., 2020). Virtual manipulatives are defined as an interactive, technology-enabled visual representation of a dynamic mathematical object, including all of the programmable features that enable it to be manipulated, that presents opportunities for constructing mathematical knowledge (Moyer-Packenham and Bolyard, 2016). They are viewed as three-dimensional objects that appear on a computer screen and can be transformed in multiple ways by the user and similar to concrete, except the images of manipulatives available through websites (i.e., internet-based) or tablets (e.g., app-based) (Bassette et al., 2019). Compared with concrete manipulatives, virtual ones can be altered randomly according to need, including changing the shape of the objects and they are more accessible because of the online environment. Anyhow, existing reviews showed both concrete and virtual manipulatives were effective for students in math learning (Bouck and Park, 2018; Peltier et al., 2020).

Furthermore, the use of manipulatives was an evidence-based effective instructional approach to teaching math to students with disabilities, including autism spectrum disorder (ASD) (Bassette et al., 2020; Bouck and Park, 2018; Peltier et al., 2020). ASD, one neurodevelopmental disorder, is characterized by persistent deficits in social communication and social interaction across multiple contexts and restricted, repetitive patterns of behavior, interests, or activities (DSM-5, 2013) and currently affects one in every 54 eight-year-olds (Maenner et al., 2020). Research showed that students with ASD often have trouble developing problem solving and critical thinking skills in math (Hua et al., 2012), which require high level thinking and comprehension of abstract concepts. Also, Whitby (2013) concluded that students with ASD demonstrated difficulties with abstract concepts. Considering the challenges which students with ASD face in the processing of abstract concepts, learning with manipulatives was a good way for them.

There is some evidence to suggest that manipulatives interventions improved the mathematical performance for ASD. For instance, Shurr et al. (2021) examined effects of manipulatives in the acquisition of double-digit addition and word problem-solving abilities of three elementary students with ASD using a single-case experimental design and the results presented that these interventions produce better outcomes than baseline. Jimenez and Besaw (2020) investigated the impact of virtual manipulatives, paired with graphic organizers and systematic instruction, for two elementary students with ASD and moderate intellectual disability to gain early numeracy skills and indicated a functional relationship between the use of virtual manipulatives and student math skills, supported by statistical analysis with a large effect. Additionally, students were able to generalize and maintain the skills across new math contexts. In study of Bassette et al. (2020) which compared the use of concrete and virtual manipulatives in teaching subtraction skills to elementary students with ASD by an alternative treatment design, all participants improved the ability after the treatment while only two of three participants demonstrated improved maintenance scores. Therefore, in general, manipulatives are effective instructional practice for ASD in math learning while the effects of manipulatives interventions in generalization and maintenance are yet to be discussed.

Earlier, Sowell (1989) carried out meta-analysis of the use of manipulatives applied to mathematics learning. In the study, each achievement and attitude effect size were estimated using a formula from Glass et al. (1981) and the results of 60 studies showed that mathematics achievement was increased through the long-term use of concrete manipulatives. One limitation of Sowell's synthesis was that it did not examine whether instructional characteristics or other factors moderate the effectiveness of manipulatives in math learning. Therefore, Carbonnean et al. (2013) performed a meta-analysis of 55 studies, including 30 studies designed of quasi-experiment, 13 experiments and 12 within subjects, to examine the efficacy of teaching mathematics with concrete manipulatives when compared to instruction with no concrete materials and to identify potential moderators including instructional and methodological characteristics. Aggregated mean effect size calculated by Cohen's *d* of 0.37 was statistically significant, in favor of the use of concrete manipulatives. Further, the effect of interventions using concrete manipulatives was moderated by both instructional (e.g., developmental status, math topic, instructional time) and methodological (e.g., peer-review status, research design, test type) characteristics of the studies.

The above two reviews all focused on typically developing children. For students with disabilities, Bouck and Park (2018) reviewed 36 articles involving mathematics manipulatives, both concrete and virtual, and summarized each study of participant characteristics, study design, mathematical content and manipulatives, effect of manipulatives and quality indicators. Of the 36, 21 were single-case design studies and

15 were group design studies. Of the 21single case studies, seven met the Horner et al. (2005) quality indicators and two of the seven studies which could be evaluated relative to the Gersten et al. (2005) experimental/quasi-experimental quality indicators were met the quality indicators, which could conclude that most of the researches that exist were low in scientific credibility or couldn't be evaluated by quality indicators. What's more, Bouck et al. (2018) collated the results of each study to analyze the impacts of manipulatives while the effect size was not calculated. Based on existing reviews and meta-analysis, Peltier et al. (2020) conducted further investigation on the effectiveness of mathematics manipulatives on students at risk or identified with a disability and explored whether the effects vary based on systematic differences related to intervention design or population characteristics. They focused on the studies with single-case experimental design, evaluating the methodological quality of studies based What Works Clearinghouse (WWC) (Kratochwill et al., 2013), measuring the effect of mathematics interventions on child outcomes using manipulatives via visual analysis, Tau-U and between-case standardized mean difference (BC-SMD), and identifying the different effects do participant characteristics, manipulative characteristics and interventions have by moderator analysis. Overall, 53 studies were met inclusion criteria and 48 studies were included in the omnibus effect size. Omnibus Tau-U effect size was 0.91 and the BC-SMD for individual studies ranged from 0.03 to 18.58, suggesting manipulatives were effective at improving the mathematical performance of students at risk or identified with a disability. Thirty-three of the 48 studies met indicators with or without reservations. Moderator analyses revealed that, of all the variables, only disability category served as a moderator.

As discussed above, they both concentrated on students with disabilities, not exclusively on ASD. What's more, the existing reviews of literature for mathematics manipulatives tend to highlight the effect of intervention, that is whether students acquire mathematical skills after intervention. However, learning for students with disabilities, including ASD, occurs in four stages: acquisition, fluency, generalization and maintenance (Alberto and Troutman, 2009; Collins et al., 2012; Shurr et al., 2019). Acquisition is the initial learning of a new behavior or response. Fluency is how well a learner can perform a specific behavior. Acquisition and fluency were most of the researchers focusing on. While, if a learner acquires a skill that does not maintain or generalize, instruction has little meaning. Generalization, perhaps the most important phase of learning, is the ability to perform a behavior across different conditions, including people, settings, activities, materials, and times of day. If learners cannot generalize or apply behaviors that have been acquired, then learning has no purpose. Maintenance refers to the ability of a learner to perform a behavior over time. In general, while students need to first acquire a skill before they can become fluent, the ultimate goal is for students to maintain the skill over time and generalize across settings, context, people, and materials (Collins et al., 2012). Each of the stages is significant for learning, while few researchers focused on skill generalization and maintenance for students with ASD. Lafay et al. (2019) conducted a systematic review to examine the immediate effects as well as maintenance and transfer of interventions with manipulatives on mathematics learning disabilities (MLD) by assessing the methodological quality. A total of 38 studies were listed, with 16 group studies and 22 single-case studies. To determine the level of methodological quality of each study, they utilized the quality indicators outlined by Gersten et al. (2005) to each group study and Horner et al. (2005) to each single-case study. The results suggested that mathematics interventions overall with manipulatives were effective for MLD. Yet, because few articles that assessed maintenance and transfer and meet the quality standards, it was unable for Lafay et al. (2019) to conclude that interventions in these studies are evidence-based practice. Lafay et al. (2019) gave a systematic review of the effects of intervention with manipulatives on immediate learning, maintenance and transfer in individuals with MLD and did not calculate the effect size or take instructional variables, student characteristics and other possible confounding and moderating variables into consideration.

Consequently, the purpose of the present study was to examine the effects of generalization and maintenance in mathematics for ASD using manipulatives. In addition, considering the necessity of individualized interventions for children with ASD, single-case experimental design can be more appropriate. Single-case experimental designs identified the influence of variables on specific behavior of a specific individual by monitoring their performance in manipulating independent variables. The performance of the monitored individual over a period of time is recorded. Individual's performance can be compared under different experimental conditions or manipulations of independent variables. As such, each individual is considered as a unit of analysis and acts as his or her own comparison (Odom et al., 2005). Meta-analysis techniques are then used to synthesize and analyze data from many single-case experiments, identifying the effect of the use of manipulatives in maintaining and/or generalizing mathematical performance of individuals with ASD by using a single metric applicable to all studies.

To sum up, we focused on the single-case studies and aimed to extend the literature by evaluating the methodological quality by WWC Design standards, reporting the effects of generalization and maintenance by calculating effect sizes; and identifying whether effects vary based on different variables related to participants characteristics, intervention design or mathematical content. This meta-analysis will help researchers to find out the current state of the extant literature and make decisions on manipulative selection and intervention design to maximize individuals' performances in generalization and maintenance. The following research questions will be solved: (a) What is the status of the extant literature regarding the measures on maintenance and/or generalization of effects of manipulatives interventions on mathematical content for participants with ASD; (b) What is the magnitude of effect (i.e., Tau-U effect sizes, non-overlap of all pairs (NAP)) of manipulatives interventions for maintaining and/or generalizing mathematical performance of individuals with ASD and (c) What effects do participant characteristics, study design, intervention characteristics and mathematical content have on maintenance and/or generalization of the effects of mathematics interventions using manipulatives?

2. Method

A comprehensive search was conducted for all studies investigating the effects of manipulatives for ASD in math study during generalization and/or maintenance phases. Search methods were consistent with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA, Moher et al., 2009), including four steps. Fig. 1 (on the next page) contains a detailed description of the search to identify eligible studies. First, an electronic search was conducted within seven electronic databases (i.e., ProQuest, ProQuest Dissertations & Theses Global, Academic Search Premier, Education Resources Information Center, PsycARTICLES, PsycINFO, and Teacher Reference Center) unlimited to the year of publication. The selection was restricted to peer-reviewed articles or dissertations published in English. For each search, the following search terms were used: (Field 1) *manipulative* with app**, OR *computer**, OR *virtual**, OR *digital**, OR *technolog**, OR *math**, OR *concrete**, OR *physical**; (Field 2) *autis**, OR *Asperger*, OR *autism spectrum disorder*, OR *ASD*, OR *PDD*, OR *pervasive developmental dis**, OR *developmental dis**, OR *DD*; (Field 3) *math**, OR *problem solving*, OR *numeracy*, OR *computation*, OR *geometry*, OR *statistic*, OR *concept*, OR *algebra*, OR *calculation*, OR *fraction*, OR *arithmetic*. The search resulted in 1668 articles and 1456 after excluding the duplications.

2.1. Inclusion criteria

To be included in the review, the following criteria for inclusion were used for eligibility. The studies (a) used a single-case design, (b) included at least one participant with ASD, (c) used a manipulative (i.e., a concrete or virtual/digital object a student would manipulate or move to aid in understanding or solving mathematics problems) as primary intervention component, (d) had at least one dependent variable relative to mathematical learning or skill acquisition, (e) collected maintenance and/or generalization data for the dependent variable relative to mathematics.

2.2. Abstract search and full text review

Applying the aforementioned inclusion criteria, a review of titles and abstracts excluded 1395 articles. If a decision could not be made upon the title and abstract alone, the article was retained for full-text screening. The full texts of each of the remaining 61 studies were screened against inclusion criteria. Two unavailable studies were excluded. Finally, a total of 17 articles were included for further analysis.

2.3. Hand search

Once all electronic files were audited and studies were chosen for inclusion in the review, the second step was to conduct a hand search within the following journals: *Exceptional Children, Journal of Special Education, Remedial and Special Education, Journal of Positive Behavior Interventions, Research in Development Disabilities, AJIDD-American Journal on Intellectual and Developmental Disabilities* and *Education and Treatment of Children*. All articles published from January 2019 to September 2021were screening for eligibility, while no more articles were included.

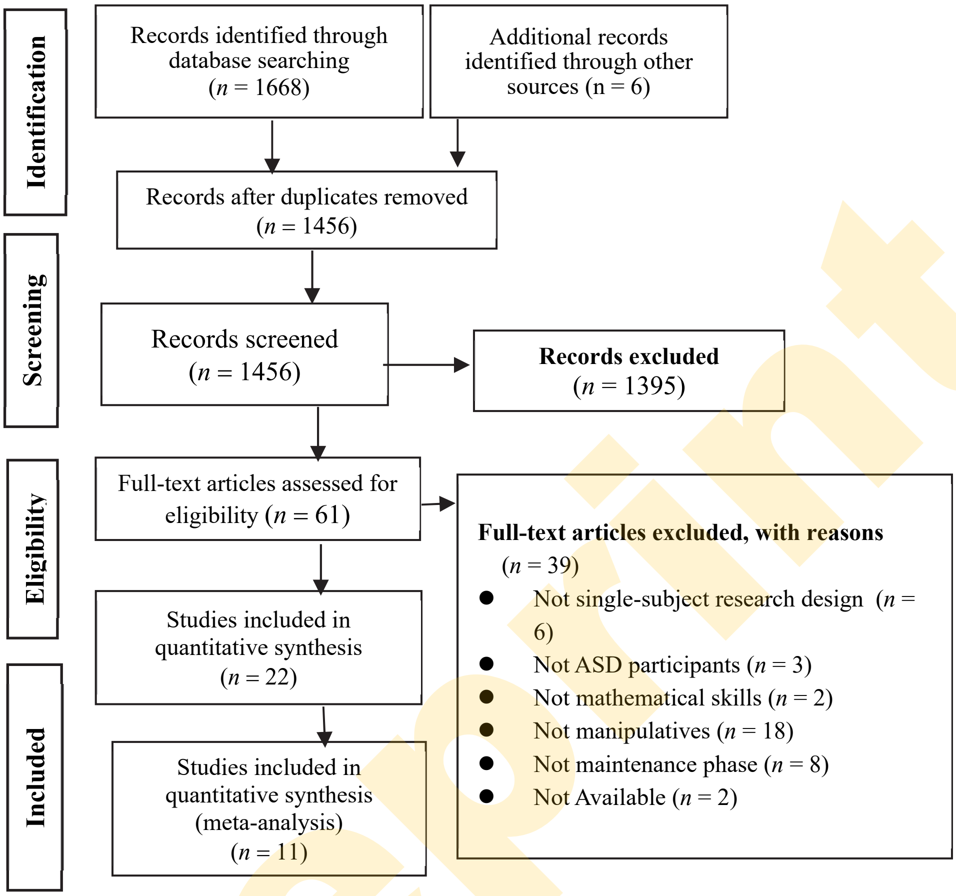

Fig. 1. Literature searches and results

2.4. *Reference and citation search*

The third step was to conduct a reference search. Five relevant meta-analysis and review articles' references were screened and yielded one article which met the inclusion criteria. Finally, to increase the likelihood that all the potentially relevant studied were identified, a citation search was conducted by reviewing all articles which had cited the included studies. Five additional studies were identified in the reviews. In total, 22 articles were included in this review for further analysis.

2.5. *Quality assessment of studies*

The methodological quality of the 22 included studies was evaluated using the WWC Pilot Single-Case Design Standards (Kratochwill et al., 2010). Design Standard 1 evaluated whether the data in the article were presented in graphical and/or tabular format. Design Standard 2 measured if the independent variable was systematically manipulated. Design Standard 3A to 3C assessed for inter-assessor agreement (IAA). Design Standard 3A evaluated whether the study reported IAA. Design Standard 3B measured if the study collected IAA in each phase and at least 20 percent of data points in each phase. Design Standard 3C assessed whether the values of IAA were at least 0.8 measured by percentage agreement or 0.6, if measured by Cohen's kappa. Design Standard 4A to 4B were about the intervention. Design Standard 4A evaluated whether the study demonstrated at least three attempts to treatment effects at least three different points in time. Design Standard 4B measured whether the study met criteria involving the number of data points depending on the design type. Design Standard 5A to 5C were additional criteria specially for multiple probe designs, as following: (a) initial pre-intervention data collection sessions must overlap vertically, (b) probe points must be available just prior to introducing the independent variable, and (c) each case not receiving the intervention must have a probe point in a session where another case either first receives the intervention or reaches the pre-specified intervention criterion. Following the application of the design standards, each article was assigned to a score for overall design classification to indicate whether the study's design "Met Design Standards", "Met Design Standards with Reservations", or "Did not Meet Standards".

2.6. Coding of studies

Referring to the work conducted by previous researchers (Bouck and Park, 2018; Carbonneau et al., 2013; Peltier et al., 2020; Spooner, 2019), the included articles that met WWC standards with or without reservations were summarized on the following categories: (1) participant characteristics, including gender, age and co-occurring diagnosis; (2) study design, (3) intervention characteristics, containing interventionist, the type of the manipulatives used (e.g., concrete, virtual) and the instructional sequences (e.g., concrete-representational-abstract (CRA), virtual-representational-abstract (VRA), virtual-abstract (VA) , and virtual-representational (VR) sequences) (Bouck et al., 2021) (4) mathematical content including number and operation, algebra, geometry, measurement, data analysis and probability of the NCTM (2000).

Furthermore, according to Schlosser and Lee (2000) and Neely et al. (2016), generalization and maintenance variables could be summarized according to the following categories: (a) generalization dimension, (b) generalization assessment design, (c) maintenance assessment design, and (d) latency to maintenance probes. The generalization dimension included three categories: (a) setting (i.e., the data was collected across setting), (b) material or behavior, (i.e., the data was collected across material or behaviors) (c) person (i.e., the data was collected across persons). The generalization assessment design contained three categories: (a) single probe (i.e., one data was collected in a generalization session), (b) multiple probes (more than one probe were collected in the duration of the study), and (c) continuous probes (i.e., the data were collected during the baseline, intervention and post-intervention session). The maintenance assessment design variable included three categories: (a) single probe (i.e., only one maintenance data point was collected in a post-intervention phase), (b) multiple probes (i.e., more than one probes were collected during the maintenance phase), and (c) sequential withdrawn (i.e., the intervention components were sequentially withdrawn in consecutive experimental phases). In addition, the description of the latency to maintenance probes was coded (i.e., 2-week follow-up). We also coded the number of the sessions during the maintenance phases.

2.7. Data extraction

Numerical values for each graphed data point in each study were extracted to format graphed data into comma separated files by a web-based tool WebPlotDigitizer (https://automeris.io/WebPlotDigitizer/). In the event that data measured skills of participants without ASD, these data were excluded from the analysis. The extracted data were categorized by baseline, intervention, generalization and maintenance phases.

2.8. Data analysis

In the review, Tau-U and NAP, measures of effect size, were calculated for each study with the web-based Tau-U and NAP calculator, available on singlecaseresearch.org (Vannest et al., 2016). Due to no consensus on which effect size measure is best for addressing the complexity of single-case studies currently, it is better to compute more than one measure when synthesizing the literature. Tau-U is interpreted as the percent of nonoverlapping data minus the percent of overlapping data (Parker et al., 2011). NAP is defined as the percentage of all pairwise comparisons across Phases A and B, which show the percentage of data which improve across phases (Parker and Vannest, 2009). In this meta-analysis, Tau-U and NAP were selected as the effect size measure because they both have greater statistical power and precision, simple calculation and the ability to calculate confidence intervals. In addition, they tend to be less susceptible to outliers (Parker et al., 2011). In the study, data in the maintenance or generalization phases were contrasted with both baseline and intervention phases within a participant/condition. For example, for one condition in a multiple-baseline design with ABC design, in which C collected maintenance or generalization data, one contrast would be A-C and a second B-C. Different resulting Tau-U or NAP scores indicate different effects. Tau-U scores less than or equal to 0.62 indicate a small effect, 0.63-0.92 a medium effect, and 0.93 and above a large effect (Parker et al., 2011). And for NAP scores, values less than or equal to 0.65 indicated a small effect, 0.66−0.92 a medium effect, and 0.93 and above a large effect (Parker and Vannest, 2009).

After Tau-U effect sizes were calculated, we enter these data into the Comprehensive Meta-Analysis software program (Version 3; Borenstein et al., 2005), along with each associated standard error (SD_{Tau}) to conduct moderator analyses. A random effects model was preferred in this case because the studies included in this meta-analysis vary in the participants, outcome measures, procedures, and settings, and it was hypothesized that the variance between studies was on account of systematic differences instead of sampling error alone (Borenstein et al., 2009; Lipsey and Wilson, 2001). Moderator analyses generating an effect size for each potential moderator and its associated subgroups and statistically significant were detected by analyzing associated p value for the between study variance (i.e., Q_b).

2.9. Inter-observer agreement

Inter-Observer Agreement (IOA) was conducted on all aspects of the searches, including the initial screening of inclusion criteria, descriptive study characteristics and data extraction, to ensure all the appropriate studies had been included and correct information had been recorded. Agreement between raters defined as both raters have determined whether to include or exclude the same study, and if both raters agreed that the information represented in the data extraction table was an accurate representation of the study. IOA scores were calculated by dividing the agreements between the two raters by agreements plus disagreements and multiplying by 100.

In the study, the first and second authors, who have got professional training, screened the articles independently. For each stage of the literature review, 50% of articles were screened by the second author. The IOA results of the two raters were as follows: 97.4% for abstract screening, 92.3% for the full text review, 100% for hand search, 100% for reference search, 97.8% for citation search. Also, IOA was conducted on all the articles which meet the inclusion criteria of the WWC Design Standard coding and the result was 99.5%. Additionally, the IOA was 97.2% for study coding and 99.3% for data extraction of all the articles which were determined to have "Met Design Standards" or "Met Design Standards with Reservations". All disagreements between two raters were discussed and resolved by consensus.

3. Results

3.1. Quality of studies

Of the 22 articles that met the pre-set inclusion criteria, 11 of 22 articles (50%) were determined to have "Met Design Standards" or "Met Design Standards with Reservations" which included two dissertations. Specially, three of the 11 articles gathered data points during generalization phases, five of them gathered data points during maintenance phases, only three of them gathered both generalization and maintenance data points.

3.2. Characteristics of studies

Tab. 1 shows the general characteristics of each of the 11 articles in which the participant characteristics, study design characteristics, intervention characteristics and mathematical content of individual studies are diverse. Tab. 2 and Tab. 3 present the tau-U and NAP scores of generalization and maintenance respectively.

Tab. 1. Descriptive of individual studies

Studies	Gender	Age	Co-occurring Diagnosis	Study Design	Interventionist	Manipulative Type	Instructional Sequences	Mathematical Content	Generalization Dimension	Generalization Assessment Design	Number of Generalization Sessions	Maintenance Assessment Design	Latency to Maintenance Probes	Number of Maintenance Sessions
Agrawal (2013)	M5 F1	E	ADHD (2) PDD-NOS (1)	MBD-P	Researcher	Concrete	CRA	Number and Operation (fraction)	Material or behavior	Multiple probes	3	Multiple probes	4 weeks in experiment 1; 2 weeks in experiment 2	2
Bassette et al. (2019)	M3	E	--	ATD	Researcher	Concrete and virtual	--	Number and Operation (subtraction)	--	--	--	Multiple probes	1-2 weeks	3
Bassette et al. (2020)	M3	E	--	ATD	Researcher	Concrete and virtual	--	Number and Operation (fraction)	--	--	--	Multiple probes	1-2 weeks	3
Bouck et al. (2019)	F1	M	--	MPD-P	Researcher	Virtual	VR	Number and Operation (fraction)	--	--	--	Multiple probes	2 weeks	2
Bouck, Park, Levy et al. (2020)	F1	M	--	MPD-P	Researcher	Virtual	--	Number and Operation (division)	Material or behavior	Multiple probes	3	--	--	--
Bouck and Park (2020)	F1	M	--	MPD-P	Researcher	Virtual	--	Number and Operation (addition)	Material or behavior	Multiple probes	5	Multiple probes	2 and 4 weeks	4
Bouck et al. (2020)	F1	M	--	MPD-P	Researcher	Virtual	VR	Number and Operation (fraction)	--	--	--	Multiple probes	2 weeks	2
Cihak and Grim (2008)	--	H (4)	ID (4)	MPD-B	Teacher	Concrete	--	Number and Operation (money)	Setting	Multiple probes	3	Single probe	6 weeks	1
Weng (2019)	M5	M (3) H (2)	ID (1)	ATD	Researcher	Virtual	--	Number and Operation (money)	setting	Multiple probes	4	--	--	--
Yakubova et al. (2016)	M3 F1	P (1) E (3)	--	MBD-B	Researcher	Concrete	CRA	Number and Operation (mixed)	--	--	--	Multiple probes	3weeks	3
Yakubova et al. (2020)	M2 F1	M (3)	ADHD/SLD/ EP (1) APD/SLD (1)	MPD-P	Researcher	Concrete	--	Number and Operation (fraction)	Material or behavior	Multiple probes/ single probe	3,3,1	--	--	--

***NOTE*.** P = preschool-aged; E = elementary-school-aged; M = middle-school-aged; H = high-school-aged; ADHD = attention deficit and hyperactivity disorder; PDD-NOS = pervasive developmental disorder not otherwise specified; ID = intellectual disabilities; SLD = specific learning disorder; EP = epilepsy; APD = auditory processing disorder; MBD-P = multiple-baseline design across participants; ATD = alternative treatment design; MPD-P = multiple-probe design across participants; MPD-B = multiple-probe design across behaviors; MBD-B = multiple-baseline design across behaviors; CRA = concrete-representational-abstract; VR = virtual-representational

Tab. 2. Tau-U and NAP effect sizes per study: generalization

Study	Baseline vs Generalization		Intervention vs Generalization		Mean Effect Size Per Study	
	Tau-U	NAP	Tau-U	NAP	Tau-U	NAP
Agrawal (2013) Experiment 1	1.00 CI$_{95}$ [0.77, 1.00]	1.00 CI$_{95}$ [0.77, 1.00]	0.51 CI$_{95}$ [0.27, 0.74]	0.75 CI$_{95}$ [0.52, 0.99]	0.76 CI$_{95}$ [0.59, 0.92]	0.88 CI$_{95}$ [0.71, 1.00]
Agrawal (2013) Experiment 2	1.00 CI$_{95}$ [0.77, 1.00]	1.00 CI$_{95}$ [0.77, 1.00]	0.61 CI$_{95}$ [0.37, 0.85]	0.81 CI$_{95}$ [0.57, 1.00]	0.81 CI$_{95}$ [0.64, 0.97]	0.90 CI$_{95}$ [0.74, 1.00]
Bouck, Park, Levy et al. (2020)	1.00 CI$_{95}$ [0.40, 1.00]	1.00 CI$_{95}$ [0.40, 1.00]	-0.47 CI$_{95}$ [-1, 0.15]	0.27 CI$_{95}$ [-0.35, 0.89]	0.28 CI$_{95}$ [-0.33, 0.89]	0.64 CI$_{95}$ [0.03, 1.00]
Bouck and Park (2020)	0.80 CI$_{95}$ [0.33, 1]	0.90 CI$_{95}$ [0.43, 1]	-0.40 CI$_{95}$ [-0.93, 0.13]	0.30 CI$_{95}$ [-0.23, 0.83]	0.23 CI$_{95}$ [-0.27, 0.74]	0.62 CI$_{95}$ [0.11, 1]
Cihak and Grim (2008)	1.00 CI$_{95}$ [0.77, 1.00]	1.00 CI$_{95}$ [0.77, 1.00]	0.48 CI$_{95}$ [0.27, 0.70]	0.74 CI$_{95}$ [0.53, 0.95]	0.73 CI$_{95}$ [0.57, 0.89]	0.87 CI$_{95}$ [0.71, 1.00]
Weng (2019)	0.92 CI$_{95}$ [0.56, 1.00]	0.96 CI$_{95}$ [0.60, 1.00]	-0.07 CI$_{95}$ [-0.29, 0.15]	0.46 CI$_{95}$ [0.24, 0.68]	0.30 CI$_{95}$ [0.09, 0.51]	0.65 CI$_{95}$ [0.44, 0.86]
Yakubova et al. (2020)	0.50 CI$_{95}$ [-0.10, 1]	0.75 CI$_{95}$ [0.15, 1.00]	-0.30 CI$_{95}$ [-0.91, 0.31]	0.35 CI$_{95}$ [-0.26,0.96]	0.20 CI$_{95}$ [-0.20, 0.60]	0.60 CI$_{95}$ [0.20, 1.00]
Mean Effect Size	0.94 CI$_{95}$ [0.75, 1.00]	0.97 CI$_{95}$ [0.78, 1.00]	0.23 CI$_{95}$ [0.04, 0.43]	0.62 CI$_{95}$ [0.42, 0.81]	0.63 CI$_{95}$ [0.55, 0.71]	0.81 CI$_{95}$ [0.73, 0.89]

Tab. 3. Tau-U and NAP effect sizes per study: maintenance

Study	Baseline vs Maintenance		Intervention vs Maintenance		Mean Effect Size Per Study	
	Tau-U	NAP	Tau-U	NAP	Tau-U	NAP
Agrawal (2013) Experiment 1	1.00 CI$_{95}$ [0.73, 1.00]	1.00 CI$_{95}$ [0.73, 1.00]	0.41 CI$_{95}$ [0.13, 0.68]	0.70 CI$_{95}$ [0.43, 0.98]	0.71 CI$_{95}$ [0.51, 0.90]	0.85 CI$_{95}$ [0.66, 1.00]
Agrawal (2013) Experiment 2	1.00 CI$_{95}$ [0.73, 1.00]	1.00 CI$_{95}$ [0.73, 1.00]	0.59 CI$_{95}$ [0.32, 0.87]	0.80 CI$_{95}$ [0.52, 1.00]	0.80 CI$_{95}$ [0.60, 0.99]	0.90 CI$_{95}$ [0.71, 1.00]
Bassette et al. (2019)	0.67 CI$_{95}$ [0.16, 1.00]	0.83 CI$_{95}$ [0.33, 1.00]	0.36 CI$_{95}$ [-0.15, 0.86]	0.68 CI$_{95}$ [0.17, 1.00]	0.51 CI$_{95}$ [0.15, 0.87]	0.76 CI$_{95}$ [0.40, 1.00]
Bassette et al. (2020)	0.00 CI$_{95}$ [-0.49, 0.49]	0.50 CI$_{95}$ [0.01, 0.99]	-0.20 CI$_{95}$ [-0.71, 0.31]	0.40 CI$_{95}$ [-0.11, 0.91]	-0.10 CI$_{95}$ [-0.45, 0.25]	0.45 CI$_{95}$ [0.10, 0.80]
Bouck et al. (2019)	1.00 CI$_{95}$ [0.28, 1.00]	1.00 CI$_{95}$ [0.28, 1.00]	0.00 CI$_{95}$ [-0.69, 0.69]	0.50 CI$_{95}$ [-0.19, 1.00]	0.49 CI$_{95}$ [-0.21, 1.00]	0.75 CI$_{95}$ [0.04, 1.00]
Bouck et al. (2020)	1.00 CI$_{95}$ [0.25, 1.00]	1.00 CI$_{95}$ [0.25, 1.00]	-0.38 CI$_{95}$ [-1.00, 0.29]	0.31 CI$_{95}$ [-0.35, 0.98]	0.27 CI$_{95}$ [-0.44, 0.98]	0.64 CI$_{95}$ [-0.07, 1.00]
Bouck and Park (2020)	1.00 CI$_{95}$ [0.49, 1.00]	1.00 CI$_{95}$ [0.49, 1.00]	-0.75 CI$_{95}$ [-1.00, -0.18]	0.13 CI$_{95}$ [-0.44, 0.69]	0.17 CI$_{95}$ [-0.37, 0.71]	0.59 CI$_{95}$ [0.05, 1.00]
Cihak and Grim (2008)	1.00 CI$_{95}$ [0.63, 1.00]	1.00 CI$_{95}$ [0.63, 1.00]	0.48 CI$_{95}$ [0.13, 0.83]	0.74 CI$_{95}$ [0.39, 1.00]	0.73 CI$_{95}$ [0.48, 0.99]	0.87 CI$_{95}$ [0.61, 1.00]
Yakubova et al. (2016)	0.78 CI$_{95}$ [0.55, 1.00]	0.89 CI$_{95}$ [0.65, 1.00]	-0.08 CI$_{95}$ [-0.30, 0.14]	0.46 CI$_{95}$ [0.24, 0.68]	0.34 CI$_{95}$ [0.18, 0.50]	0.67 CI$_{95}$ [0.51, 0.83]
Mean Effect Size	0.83 CI$_{95}$ [0.62, 1.00]	0.92 CI$_{95}$ [0.71, 1.00]	0.16 CI$_{95}$ [-0.04, 0.37]	0.58 CI$_{95}$ [0.38, 0.78]	0.53 CI$_{95}$ [0.45, 0.62]	0.77 CI$_{95}$ [0.68, 0.85]

3.2.1. Participant characteristics

Of all studies, there were 32 participants. Excluding four subjects whose gender was not specified in the study, 21 of the 28 participants were male (75%). For the age, 15 participants (46.88%) were elementary-aged students, 10 (31.25%) were middle school students, six (18.75%) were high school students and only one (3.13%) was a preschool student. Besides ASD, five participants (15.63%) had comorbid disabilities of intellectual disabilities, two (6.25%) with co-occurring attention deficit and hyperactivity disorder and one (3.13%) with pervasive developmental disorder not otherwise specified. What's more, one ASD case was diagnosed as specific learning disorder (SLD), and the other one was diagnosed as comorbid auditory processing disorder and SLD.

3.2.2. Study design characteristics

Among all the studies, most of the articles ($\boldsymbol{n}$ = 6, 54.55%) were designed of multiple probe design, in which, a multiple probe across participants design of single-case research design was used in five studies and the remained used a multiple probe across behavior design. Besides, two (18.18%) articles were designed of multiple-baseline design including multiple-baseline across participants and multiple-baseline across behavior. The remaining three articles (27.27%) were designed of alternative treatment design.

3.2.3. Intervention characteristics

Of the eleven studies, ten studies' interventionists were researcher (90.91%), and only one study's interventionist was teacher (9.91%). For the type of the manipulative used for intervention, four of the eleven studies (36.36%) employed concrete manipulative (e.g., base 10 blocks, colored chips, flashcards), five (45.45%) used virtual manipulatives (e.g., Fraction Tiles app, Number Lines app), and two (18.18%) conducted both concrete and virtual manipulatives. Also, two studies (18.18%) used the CRA framework and two (18.18%) studies used the virtual-representation (VR) framework.

3.2.4. Mathematical content

All of the studies focused on the number and operation. Furthermore, five studies (45.45%) addressed fraction problems; four (36.36%) focused on the basic operations, such as the subtraction; and two (18.18%) were about money.

3.2.5. Maintenance characteristics

Across the eight articles that collected maintenance of intervention effects, seven collected multiple maintenance data points (87.5%) and only one collected single maintenance follow-up data point (12.5%). None collected maintenance data using a sequential withdrawal design.

For all articles, data was collected anywhere from one week following completion of the intervention phase to six weeks following the intervention. Six articles (75%) collected maintenance data within four weeks after the conclusion of the intervention phase. One article (12.5%) collected data up to six months following the intervention phase. Moreover, one article conducted two experiments, whose follow-up probes were harvested two weeks in experiment one and four weeks in experiment two.

Finally, half of the eight studies extracted at least three data points within each maintenance phase.

3.2.6. Generalization characteristics

Of the six studies which collected data point during generalization, five of them collected multiple probes in the duration of the study. The remaining one study collected only one probe for one of the subjects and multiple probes for the other subjects. Additionally, two of the six studies assessed generalization of effects across settings, and four evaluated across behavior and materials.

3.3. Overall effect size

Overall, raw data for a total of 211 separate contracts (i.e., baseline/intervention vs. maintenance/generalization) from 11 articles with 32 participants were extracted to calculate effect sizes. Tab. 2 and 3 present the results of Tau-U and NAP scores across articles. Results from Baseline vs. Maintenance comparisons were medium (mean NAP = 0.92 and mean Tau-U = 0.83) with a variable range of effect sizes (NAP = 0.50~1.00 and Tau-U = 0.00~1.00). Intervention vs. Maintenance comparisons produced small findings with a mean NAP of 0.58 (0.13~0.80) and Tau-U of 0.16 (−0.75~0.59). Results from Baseline vs. Generalization comparison were significant with a mean NAP of 0.97 (1.00~0.75) and Tau-U = 0.94 (0.50~1.00). Intervention vs. Generalization comparisons were small with a mean NAP of 0.62 (0.27~0.81) and Tau-U = 0.23 (−0.47~0.61).

3.4. Moderator analysis

To identify whether effects with respect to generalization and maintenance varied across participant characteristics, study design, intervention characteristics and mathematical content, seven variables were examined: age, gender, co-occurring diagnosis, study design, interventionist, the type of manipulatives, mathematical content. Tab. 4 and 5 summarize the results from the analysis of the moderator analysis.

Firstly, for baseline to maintenance in comparison, of all the variables analyzed, the type of manipulative variable was the only variable that had statistically significant differences between the categories analyzed ($\boldsymbol{Q = 6.64,\ p = 0.04}$). And the mean effect size for studies with virtual manipulative was statistically greater than that of concrete or virtual/concrete manipulative. The participant characteristics variables, including gender, age and co-occurring diagnosis, did not function as moderators. Meanwhile, study design, interventionist variable and mathematical content did not find statistically significant.

Whereas, for baseline to generalization comparison, all of the variables, including participant characteristics, study design, intervention characteristics and mathematical content, did not show statistical differences.

4. Discussion

Through a complete screening process, 22 studies examining mathematics manipulatives interventions for individuals with the diagnosis of ASD meet the inclusion criteria, among which 11 studies met WWC standards. This meta-analytic review examined 11 studies aiming to analyze varying study characteristics, to evaluate the extent to which intervention using manipulatives for individuals with ASD contributed to generalization and maintenance in mathematics, whether the effect vary with different participant characteristics, study design, intervention characteristics and mathematical content, and to provide suggestions for practice and future research.

4.1. Major findings and implications

4.1.1. Quality of Evidence

In analyzing the 22 studies, we first evaluated the quality of evidence of the studies. We found 11 of them have met the WWC design standards with or without reservations. Since most of the studies were designed of multiple probe design, failing to meet the additional criteria specially for multiple probe designs was the primary reason that studies did not meet the WWC design standards. The same failure of additional criteria specially for multiple probe designs has also been notes in the study on using mathematics manipulatives with students at risk or identified with a disability (Peltier, 2020), indicating that the studies using multiple probe design needs to design the experimental process more carefully, and the experimental data should be collected and recorded reasonably in future studies. Additionally, failing to meet the design standards for IAA and insufficiency of data points in each phase (i.e., fewer than three data points in a phase) were also one of the reasons why the study did not meet WWC design standards, suggesting that the data integrity will need to be monitored more carefully in future studies. The poor experimental design may affect the credibility of the results. The effect size of these studies which did not meet the WWC design standards requires more careful interpretation. Therefore, the calculation of effect sizes and a moderator analysis in this study focuses only on studies which have met the WWC design standards with or without reservations.

4.1.2. Participant and intervention characteristics

We examined the study characteristics before analyzing the magnitude of effects of the mathematics manipulatives interventions. The results indicated that school-age children have been the main focus of manipulative studies on individuals with ASD. This may show a lack of knowledge about the effectiveness of manipulatives interventions for young children with ASD. In order to address the disparity between study populations, more research is needed on manipulatives interventions maintaining and/or generalizing relevant mathematics skills.

In examining the types of manipulatives interventions, we found that more studies paid attention to virtual manipulatives, which was similar to the finding in previous research (Bouck and Park, 2018). Additionally, two studies used the CRA instructional sequence and two adopted the VR instructional sequence. Based on the application of the indicators and standards, Bouck, Satsangi, et al. (2018) confirmed that the CRA instructional sequence was an evidence-based practice for students with learning disabilities. Students with ASD may benefit from this instructional sequence as well, and this assumption could to be verified in future studies. Meanwhile, with the development of virtual manipulatives, exploring the effect of VR framework is also a direction for further studies.

4.1.3. Mathematics content

We noted that target mathematical topics of studies included all focused on the number and operation, especially fraction problems and basic operations. On one hand, fractions and operations are important basic knowledge and skills of mathematics. Fraction problems are generally considered as the foundation of learning algebra and more advanced mathematics (Fuchs et al., 2014). Meanwhile, students often have difficulty learning fraction knowledge. Many middle and high school students are still unable to master the ideas and procedures taught about fractions in the elementary grades (Ni, 2001). Besides, operation is an important component of solving mathematical word problems (Fuchs and Fuchs 2002). Without computational accuracy and fluency, students would not be able to engage in higher level problem-solving skills, let alone actively participate in inclusive general education classrooms (Butler et al., 2001). NCTM (2000) even listed fluent computation as a goal for mathematics instruction. On the other hand, using manipulatives is beneficial to the instruction of number and operation. First of all, the number and operation are considered procedural skills or procedural understanding (Rittle-Johnson, 2017), which is the ability to both know which procedure to follow and complete the appropriate steps to arrive at the correct answer. Besides, there were very mature manipulatives to teach number and operation, such as Base 10 Blocks for teaching computation, fraction pieces of the fraction and flashcards for money-related skills. Interactions with manipulatives may help them better understand the knowledge. Moreover, several studies have identified manipulatives as an effective strategy for students with ASD (Bassette et al., 2020) and usually designed instruction with steps by steps (Shin et al., 2017). Therefore, for ASD, using manipulatives appears to be an effective way to teach number and operation.

However, for other mathematical content such as measurement, algebra, geometry, statistics and probability, more study is needed to evaluate the effectiveness of manipulatives. We point out the omission as a suggestion for further research.

4.1.4. Generalization and maintenance characteristics

As showed in the literature review, lack of articles involved the effects of manipulatives interventions for participants with ASD on maintaining and/or generalizing the related math skills which showed the neglect of generalization and maintenance. Notably, three of the 11 studies collected data points during generalization phases, five of them collected data points during maintenance phases, only three of them gathered both generalization and maintenance data. This is due to the fact that generalization and maintenance measures are somewhat more difficult to implement. As noted, one way to ensure that learners are maintaining what they have been taught is to conduct periodic probes over time where learners are required to perform targeted skills, requiring long-term follow-up of participants. However, in the real experiment, it may be impossible to track the subjects due to the school holidays or personal factors of participants, and therefore not enough maintenance probes can be collected (e.g., Saunders, 2014). Let alone generalization which needs to be measured across individuals, materials or settings, putting forward higher requirements for researchers. In spite of this, it suggested that practitioners and researchers should pay more attention to the generalization and maintenance of skills, which are the core segments in learning, despite the difficulties in measuring or collecting the performance of students during generalization and maintenance.

Additionally, inclusion of maintenance and/or generalization phases in quality indicators has not been identified as a requirement for methodological soundness, resulting in a lack of focus on maintenance and/or generalization phases in single-case study designs (Kratochwill et al., 2013). As Collins (2012) pointed, however, learning to do a skill in one context with one instructor did not necessarily mean individuals with significant disabilities would apply that skill (i.e., maintain and generalize it) whenever and wherever it was needed or would be useful. Thus, we suggest that future studies pay more attention to implementing the generalization and maintenance phases.

Moreover, included articles collected mean 2.5 probes during maintenance phases and 3.5 probes during generalization phases. Interpretation of single-case research data depends on the trend and slope of a data path (Kratochwill et al., 2013). A minimum of three data points is necessary to meet basic design standards, with more data points leading to stronger conclusions regarding the data set. Although half of the studies collected less than three probes during maintenance or generalization, most collected at least two probes which were a strength of the literature base. On the other hand, however, all of the studies collected maintenance data less than six weeks following completion of the intervention. Given the latency of the maintenance probes and these short follow-up time, whether the effects of these interventions can be sustained in a long run after the training period is questionable.

4.1.5. Magnitude of effects on generalization and maintenance

We especially focused on the magnitude of manipulatives intervention effects on maintaining and/or generalizing the mathematical performance of individuals with ASD. In general, it is gratifying that the use of manipulatives really improved mathematical performance of students with ASD during generalization and maintenance probes as compared to baseline probes as both Tau-U and NAP scores were positive in most of these studies.

Data from the present study provided information on the change from baseline to generalization and maintenance phases and the change from intervention to generalization and maintenance sessions. As mentioned above, the purpose of mathematics manipulative intervention included grasping the mathematical content across time and conditions as compared to baseline. Omnibus effect size for change from baseline to generalization and maintenance indicated the desired change in mathematical performance. For intervention to generalization and maintenance, the omnibus effect size indicated the slight effect. While the effect size for the change from intervention to maintenance was slightly above zero. Through the intervention, the participants did gain the targeted content and could generalize them across the conditions. However, as time went on, they may have forgotten some of these.

Meanwhile, we also concerned whether the effects on generalization and maintenance of manipulatives in math vary among participant characteristics, study design, intervention characteristics and mathematical content. Through the moderator analysis, for both generalization and maintenance, the effects regarding the effectiveness of manipulatives for supporting mathematics instruction were consistent across the participants included. This was positive because the results presented manipulatives were effective instructional methods for all students with ASD regardless of the age, gender and co-occurring diagnosis. Of note, effects were also consistent across implementer (i.e., teacher vs. researcher), which is consistent with findings from Peltier et al. (2020). This is promising because the findings suggest, with training, teachers can implement the intervention and yield comparable effects as researchers with expertise in the intervention. We thus call for the relevant training of teachers in the use of manipulatives.

The effectiveness of manipulatives for supporting mathematics instruction was consistent across a variety of mathematical content (e.g., addition, subtraction, division, mixed operation, fraction and money). However, as pointed out before, the target mathematical topics of the included studies all focus on numbers and operations and thus lack topics such as measurement, data analysis and probability, algebra and geometry. Consequently, this conclusion may not be representative and should be interpreted prudently.

Effects regarding the effectiveness of manipulatives for maintaining mathematical performance were significantly different among various types of manipulatives. The virtual manipulatives yielded larger effects than the concrete ones, and even better than the combination of virtual and concrete manipulatives. Many researchers investigated the potential of using virtual manipulatives in math learning for individuals with ASD recently and suggested that, comparing to the concrete manipulatives, students preferred virtual manipulatives (Bassette et al., 2019; Bassette et al., 2020; Bouck et al., 2014). Additionally, systematic reviews and meta-analysis which suggested that virtual manipulatives were more effective for students with disabilities in skill acquisition, comparing with concrete manipulatives (Bouck et al., 2018; Peltier et al., 2020). The participants with better skill acquisition may perform better over time. Since it is difficult for children with ASD to understand abstract concepts, teachers should try to use manipulatives, especially virtual manipulatives, in classroom teaching to help students learn, maintain and generalize mathematical concepts.

The other explanation for this result may be that children with ASD are individuals with high visual abilities (Fossett, 2004) and it is difficult for them to stay focus in one lesson (Bai et al., 2015). Virtual manipulative is a digital interactive experience that depict mathematical concepts and is one form of visual stimulation. For children with ASD, they may be an effective source of help to be applied in their learning style. Meanwhile, one of the features of virtual manipulatives was that they could focus students' attention on particular aspects of mathematical objects — aspects that they otherwise may not have paid attention to (Anderson-Pence, 2017). In addition, Suh and Moyer (2008) pointed that the concrete manipulatives may distract students' thought process and increase cognitive load. While virtual manipulatives can not only compensate for the deficiencies, but also provide additional visual information which was not available with the concrete manipulatives. Students also can be provided with individualized scaffolds through virtual manipulatives. What's more, according to Reimer and Moyer (2005), one advantage of virtual manipulatives is the capability of connecting dynamic visual images with abstract symbols, a limitation of regular manipulatives. Thus, teachers could try to make more use of virtual manipulatives in the classroom for better learning effect. Furthermore, as an example of assistive technology, virtual manipulatives can support the mathematics learning of school-age students in online and blended learning environments.

However, researchers should seek to systematically compare the implementation of concrete and virtual manipulatives for students with ASD. What's more, key details about the teacher's (or researcher's) practice, such as what he or she says to children and shows them at key moments in teaching, are often omitted. Thus, additional research on the specific process of applying virtual manipulatives to improve mathematics performance of students with ASD is still needed.

4.2. Limitations and Future Studies

While these findings suggested that manipulatives were effective in maintaining and/or generalizing mathematical skills for students who were identified as ASD, there were some limiting factors to consider when evaluating the results.

First, the review may be impacted by potential publication bias. Only studies published in peer-reviewed journals were included. Second, another limitation of this meta-analysis is related to the number of studies ($n = 11$) that were included. Only eight studies collected data points during maintenance phases and six collected generalization data. Due to the small number of studies, the analysis was limited to examining a few moderating variables. Third, effect sizes for individual studies were based on the data presented in the articles, which were extremely limited in some cases. Many of the articles presented limited generalization and maintenance data, with as few as one data point in some cases. Due to the limited number of data points used to calculate individual effect sizes, the effect sizes may have been influenced by typical variability rather than changes in the independent variable. A fourth limitation of the current study is about the outcome. The results are of limited generality because this analysis only included studies with single-case design which meet the WWC design standards with or without reservations and all included participants were identified with ASD.

As a consequence, firstly, it was necessary for future researchers to pay more attention to implementing the generalization and maintenance phases, and conduct more high-quality studies to examine the effectiveness of manipulatives in maintaining and/or generalizing mathematical skills for students identified with ASD. Meanwhile, researchers might collect more probes during generalization and maintenance phases so that present stronger evaluations of generalization and maintenance. Secondly, through the literature review of the 11 studies, this analysis has identified that the target mathematical content was limited to the number and operation. Future work can further investigate the effectiveness of manipulatives in the instruction of measurement, algebra, geometry, statistics and probability. At the same time, based rich literature, the applicability of different kinds of manipulatives to different mathematical contents may be further explored, namely determining what types of content can be effectively taught using concrete or virtual manipulatives and what type of content is difficult. Thirdly, as mentioned above, the study has found that manipulatives were effective instructional approach for maintaining and generalizing mathematical content. Future work can investigate how to maximize the potential learning benefits of concrete and virtual manipulatives and try to design effective instruction. Furthermore, as mentioned earlier, children with ASD may benefit from the CRA instructional sequence, but this finding needs to be validated by more studies. In addition, with the development of virtual manipulatives, how to apply it to improve the mathematics performance of students with ASD has become an urgent problem to be solved. At the same time, additional research on exploring the effects of VR framework is also needed and still emerging. Last but not least, a related direction for future work is identifying other factors which may be influential in the instructional effect of manipulatives but which have not been accounted in the above analysis such as factors that were not identified due to the limited literature available. Related to the virtual manipulatives, exploring the effect of VR framework is also a direction for further study. Besides, more specific information should be sought in future reviews.

Reference

*J. Agrawal (2013). *The effects of explicit instruction with manipulatives on the fraction skills of students with autism* [Unpublished doctoral dissertation]. George Mason University.

P. Alberto and A. Troutman (2009). Applied behavior analysis for teachers (8th ed.). Pearson.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

K. L. Anderson-Pence (2017). Techno-mathematical discourse: A conceptual framework for analyzing classroom discussions. *Education Sciences,* 7(1), 40. https://doi.org/10.3390/educsci7010040

Z. Bai, A. F. Blackwell, and G. Coulouris (2015). Using augmented reality to elicit pretend play for children with autism. *IEEE Transactions on Visualization and Computer Graphics,* 21(5), 598–610. https://doi.org/10.1109/TVCG.2014.2385092

*L. Bassette, E. Bouck, J. Shurr, J. Park, and M. Cremeans (2019). Comparison of Concrete and App-Based Manipulatives to Teach Subtraction Skills to Elementary Students with Autism. *Education and Training in Autism and Developmental Disabilities,* 54(4), 391–405.

*L. Bassette, E. Bouck, J. Shurr, J. Park, M. Cremeans, E. Rork, K. Miller, and S. Geiser, (2020). A comparison of manipulative use on mathematics efficiency in elementary students with autism spectrum disorder. *Journal of Special Education Technology,* 35(4), 179–190. https://doi.org/10.1177/0162643419854504

M. Borenstein, L. V. Hedges, J. Higgins, and H. Rothstein (2005). *Comprehensive Meta-Analysis 2.0*. Englewood, NJ: Biostat.

M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein (2009). *Introduction to Meta-Analysis*. West Sussex, UK: John Wiley.

E. C. Bouck, R. D. Anderson, H. Long, and J. Sprick (2021). Manipulative-based instructional sequences in mathematics for students with disabilities. *Teaching Exceptional Children*, 004005992199459. https://doi.org/10.1177/ 0040059921 9945 99

E. C. Bouck and S. M. Flanagan (2010). Virtual manipulatives: What they are and how teachers can use them. *Intervention in School and Clinic,* 45(3), 186–191. https://doi.org/10.1177/1053451209349530

*E. C. Bouck, C. Maher, J. Park, and A. Whorley (2020). Learning fractions with a virtual manipulative based graduated instructional sequence. *Education and Training in Autism and Developmental Disabilities,* 55(1), 45–59.

E. C. Bouck and J. Park (2018). A systematic review of the literature on mathematics manipulatives to support students with disabilities. *Education & Treatment of Children,* 41(1), 65–106. https://doi.org/10.1353/etc.2018.0003

*E. C. Bouck and J. Park (2020). App-based manipulatives and the system of least prompts to support acquisition, maintenance, and generalization of adding integers. *Education and Training in Autism and Developmental Disabilities,* 55(2), 158–172.

*E. C. Bouck, J. Park, K. Levy, K. Cwiakala, and A. Whorley (2020). App-based manipulatives and explicit instruction to support division with remainders. *Exceptionality*, 28(1), 45–59. https://doi.org/10.1080/09362835.2019.1586709

*E. C. Bouck, J. Park, C. Maher, K. Levy, and K. Cwiakala (2018;2019;). Acquiring the skill of identifying fractions through the virtual-abstract framework. *Journal of Developmental and Physical Disabilities,* 31(4), 435–452. https://doi.org/10.1007/s10882-018-9650-9

E. C. Bouck, R. Satsangi, and J. Park (2018). The Concrete–Representational–Abstract approach for students with learning disabilities: An evidence-based practice synthesis. *Remedial and Special Education,* 39(4), 211–228. https://doi.org/10.1177/0741932517721712

E. C. Bouck, J. Shurr, L. Bassette, J. Park, and A. Whorley (2018). Adding it up: Comparing concrete and app-based manipulatives to support students with disabilities with adding fractions. *Journal of Special Education Technology,* 33(3), 194–206. https://doi.org/10.1177/0162643418759341

F. M. Butler, S. P. Miller, K. H. Lee, and T. Pierce (2001). Teaching mathematics to students with mild-to-moderate mental retardation: A review of the literature. *Mental retardation*, 39(1), 20–31.

K. J. Carbonneau, S. C. Marley, and J. P. Selig (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology,* 105(2), 380–400. https://doi.org/10.1037/a0031084

*D. F. Cihak and J. Grim (2008). Teaching students with autism spectrum disorder and moderate intellectual disabilities to use counting-on strategies to enhance independent purchasing skills. *Research in Autism Spectrum Disorders,* 2(4), 716–727. https://doi.org/10.1016/j.rasd.2008.02.006

B. C. Collins (2012). *Systematic Instruction for Students with Moderate and Severe Disabilities*. Brookes Publishing Company.

B. Fossett (2004). Visual support strategies for literacy development. *SEA Crosscurrents Magazine*, 23–32.

L. S. Fuchs and D. Fuchs (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities*, 35(6), 564–574. https://doi.org/10.1177/0022219402 0350060701

L. S. Fuchs, R. F. Schumacher, S. K. Sterba, J. Long, J. Namkung, A. Malone, A, C. L. Hamlett, N. C. Jordan, R. Gersten, R. S. Siegler, and P. Changas (2014). Does working memory moderate the effects of fraction intervention? an aptitude-treatment interaction. *Journal of Educational Psychology,* 106(2), 499–514. https://doi.org/10.1037/a0034341

R. Gersten, L. S. Fuchs, D. Compton, M. Coyne, C. Greenwood, and M. S. Innocenti (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children,* 71(2), 149–164. https://doi.org/10.1177/001440290507100202

G. V. Glass, B. McGaw, and M. L. Smith (1981). *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage.

R. H. Horner, E. G. Carr, J. Halle, G. McGee, S. Odom, and M. Wolery (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

Y. Hua, B. S. T. Morgan, E. R. Kaldenberg, and M. Goo (2012). Cognitive strategy instruction for functional mathematical skill: Effects for young adults with intellectual disability. *Education and Training in Autism and Developmental Disabilities,* 47(3), 345–358.

B. A. Jimenez and J. Besaw (2020). Building early numeracy through virtual manipulatives for students with intellectual disability and autism. *Education and Training in Autism and Developmental Disabilities,* 55(1), 28–44.

T. R. Kratochwill, J. Hitchcock, R. H. Horner, J. R. Levin, S. L. Odom, D. M. Rindskopf, and W. R. Shadish (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/Document/229

T. R. Kratochwill, J. H. Hitchcock, R. H. Horner, J. R. Levin, S. L. Odom, D. M. Rindskopf, and W. R. Shadish (2013). Single-case intervention research design standards. *Remedial and Special Education,* 34(1), 26–38. https://doi.org/10.1177/0741932512452794

A. Lafay, H. P. Osana, and M. Valat (2019). Effects of interventions with manipulatives on immediate learning, maintenance, and transfer in children with mathematics learning disabilities: A systematic review. *Education Research International,* 2019, 1–21. https://doi.org/10.1155/2019/2142948

M. W. Lipsey and D. B. Wilson (2001). *Applied Social Research Methods Series: Vol. 49. Practical Meta-Analysis* (L. Bickman and D. J. Rog, Eds.). Thousand Oaks, CA: SAGE.

M. J. Maenner, K. A. Shaw, J. Baio, A. Washington, M. Patrick, M. DiRienzo, D. L. Christensen, L. D. Wiggins, S. Pettygrove, J. G. Andrews, M. Lopez, A. Hudson, T. Baroud, Y. Schwenk, T. White, C. R. Rosenberg, Li-Ching Lee, R. A. Harrington, M.

Huston, and A. Hewitt (2020). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, united states, 2016. *MMWR. Surveillance Summaries*, 69(4), 1–12. https://doi.org/10.15585/mmwr.ss6904a1

D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and Prisma Group. (2009). Reprint — preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. *Physical Therapy,* 89(9), 873880. https://doi.org/10.1093/ptj/89.9.873

P. S. Moyer (2001). Are we having fun yet? how teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics,* 47(2), 175–197. https://doi.org/10.1023/A:1014596316942

P. S. Moyer, J. J. Bolyard, and M. A. Spikell (2002). What are virtual manipulatives? *Teaching Children Mathematics,* 8(6), 372–377.

P. S. Moyer-Packenham and J. J. Bolyard (2016). Revisiting the definition of a virtual manipulative. In *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 3–23). Springer. https://doi.org/10.1007/978-3-319-32718-1_1

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

L. C. Neely, J. B. Ganz, J. L. Davis, M. B. Boles, E. R. Hong, J. Ninci, J, and W. D. Gilliland (2016). Generalization and maintenance of functional living skills for individuals with autism spectrum disorder: A review and meta-analysis. *Review Journal of Autism and Developmental Disorders,* 3(1), 37–47. https://doi.org/10.1007/s40489-015-0064-7

Y. Ni (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemporary Educational Psychology,* 26(3), 400–417. https://doi.org/10.1006/ceps.2000.1072

S. L. Odom, E. Brantlinger, R. Gersten, R. H. Horner, B. Thompson, and K. R. Harris (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children,* 71(2), 13–148. https://doi.org/10.1177/001440290507100201

R. I. Parker and K. Vannest (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy,* 40(4), 357–367. https://doi.org/10.1016/j.beth.2008.10.006

R. I. Parker, K. J. Vannest, and J. L. Davis (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303–322. https://doi.org/10.1177/0145445511399147

R. I. Parker, K. J. Vannest, J. L. Davis, and S. B. Sauber (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

C. Peltier, K. L. Morin, E. C. Bouck, M. E. Lingo, J. M. Pulos, F. A. Scheffler, A. Suk, L. A. Mathews, T. E. Sinclair, and M. E. Deardorff (2020). A meta-analysis of single-case research using mathematics manipulatives with students at risk or identified with a disability. *The Journal of Special Education,* 54(1), 3–15. https://doi.org/10.1177/0022466919844516

K. Reimer and P. S. Moyer (2005). Third-graders learn about fractions using virtual manipulatives: A classroom study. *The Journal of Computers in Mathematics and Science Teaching,* 24(1), 5.

B. Rittle-Johnson (2017). Developing mathematics knowledge. *Child Development Perspectives*, 11(3), 184–190. https://doi.org/10.1111/cdep.12229

A. F. Saunders (2014). *Effects of schema-based instruction delivered through computer-based video instruction on mathematical word problem solving of students with autism* *spectrum disorder and moderate intellectual disability* (Doctoral dissertation, The University of North Carolina at Charlotte).

R. W. Schlosser and D. L. Lee (2000). Promoting generalization and maintenance in augmentative and alternative communication: A meta-analysis of 20 years of effectiveness research. *Augmentative and Alternative Communication,* 16(4), 208–226. https://doi.org/10.1080/07434610012331279074

M. Shin, D. P. Bryant, B. R. Bryant, J. W. McKenna, F. Hou, and M. W. Ok (2017). Virtual manipulatives: Tools for teaching mathematics to students with learning disabilities. *Intervention in School and Clinic*, 52(3), 148–153. https://doi.org/10.1177/1053451216644830

J. Shurr, E. C. Bouck, L. Bassette, and J. Park (2021). Virtual versus concrete: A comparison of mathematics manipulatives for three elementary students with autism. *Focus on Autism and Other Developmental Disabilities,* 36(2), 71–82. https://doi.org/10.1177/1088357620986944

J. Shurr, B. Jimenez, and E. C. Bouck (2019). Educating students with intellectual disability and autism spectrum disorder: Book 1 Research-based practices and education science. *Council for Exceptional Children*.

E. Sowell (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education,* 20(5), 498–505. https://doi.org/10.2307/749423

F. Spooner, J. R. Root, A. F. Saunders, and D. M. Browder (2019). *An updated evidence-based practice review on teaching mathematics to students with moderate and severe developmental disabilities*. SAGE Publications. https://doi.org/10.1177/0741932517751055

J. M. Suh and P. S. Moyer (2008). Scaffolding special needs students' learning of fraction equivalence using virtual manipulatives. *Proceedings of the International Group for the Psychology of Mathematics Education*, 4, 297–304.

K. J. Vannest, R. I. Parker, O. Gonen, and T. Adiguezel (2016). *Single case research: Web based calculators for SCR analysis* (Version 2.0) [Web-based application]. Texas A&M University. http://www.singlecaseresearch.org/

*P. Weng and E. C. Bouck (2019). Comparing the effectiveness of two app-based number lines to teach price comparison to students with autism spectrum disorders. *Disability and Rehabilitation: Assistive Technology,* 14(3), 281–291. https://doi.org/10.1080/17483107.2018.1430869

P. J. Whitby (2013). The effects of solve it! on the mathematical word problem solving ability of adolescents with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities,* 28(2), 7–88. https://doi.org/10.1177/1088357612468764

*G. Yakubova, E. M. Hughes, and B. B. Chen (2020). Teaching students with ASD to solve fraction computations using a video modeling instructional package. *Research in Developmental Disabilities, 101*, 103637–103637. https://doi.org/10.1016/j.ridd.2020.103637

*G. Yakubova, E. M. Hughes, and M. Shinaberry (2016). Learning with technology: Video modeling with Concrete–Representational — Abstract sequencing for students with autism spectrum disorder. *Journal of Autism and Developmental Disorders,* 46(7), 2349–2362.https://doi.org/10.1007/s10803-016-2768-7