

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPROVING INDIVIDUALIZED EDUCATION PROGRAMS WITH BEHAVIOR SKILLS

TRAINING: BALANCING QUALITY AND EFFICIENCY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

JASMINE ARAE JUSTUS

Norman, Oklahoma

2025

IMPROVING INDIVIDUALIZED EDUCATION PROGRAMS WITH BEHAVIOR SKILLS
TRAINING: BALANCING QUALITY AND EFFICIENCY

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Brittany Hott, Chair

Dr. Corey Peltier

Dr. Howard Crowson

Dr. Kyong Kwon

©Copyright by JASMINE ARAE JUSTUS 2025
All Rights Reserved.

Dedication

To my family, for your unwavering patience and understanding as I navigated this daring journey. We have reached the end, and this triumph is dedicated to you.

Acknowledgements

I owe this entire process, as well as every success I have had, to my family. More than anything, I want to dedicate this dissertation to my mother, La'Sumer Jolly, and my little sister, Morgan Morphis. Thank you both for believing in me. Mom, I can never thank you enough for the life and stability that you provided. You have always supported me in ways that no one can. Morgan, you are the world's best little sister. I would never have been in this position without you and can't imagine what this journey would have been like without your support. Thank you for bringing me back to Earth every time I begin floating off.

Meme and Papa, you are the foundation of this family and have supported me throughout all my endeavors; without you both, our family wouldn't be whole. Melissa, the world's best Aunt Mo, thank you for guiding me into education; without it, I would not be where I am today. Ed, to the best stepdad a girl could hope for, thank you for choosing us. Thank you for listening to my public-school political banter and for showing me that the best way to make change is to take action.

Brandon, I know you don't want the compliment, but I wouldn't have been able to complete this process without you. Thank you for everything you did to support me, whether it was weekly bouquets of flowers, taking care of the dogs, listening to my rants about special education law, or reminding me every day that I can do hard things. I hope you know how much you mean to me.

To my fellow doctoral students, thank you for accompanying me through this process. Sarah, we started this together, and I cannot believe we're at the finish line. You held me up through some of the most difficult times in my life, because, of course, the world likes to fall apart as soon as you start a PhD program. Kevin and Julie, you both mean the world to me. I didn't know that

this journey would provide me with the best two friends anyone could ask for. Thank you both for just being you.

Dr. Hott, none of this would have been possible without you. Thank you for taking a chance on me. Your mentorship has been invaluable, and it has helped me grow into the researcher and practitioner I am today. I am deeply grateful for the time and effort you invested in my education. Dr. Crowson, Kwon, and Peltier, thank you all for being a part of my committee. I know it must have been difficult to listen to someone rant about IEP quality for four years, but your support and guidance are greatly appreciated.

Lastly, I would like to thank the school staff, students, families, and colleagues who have been there every step of the way. Thank you for allowing me to be a part of your lives. Without you all, I would never have made the move to special education. I am beyond thankful for the lessons I get to learn every day because those in my life trust me to advocate for and support students with disabilities. Our work is not done. We will continue to advocate for the services and supports that are needed. We will continue to do what is best for all students. The world is a better place with you all in it, please remember that.

Table of Contents

LIST OF TABLES	xi
LIST OF FIGURES	xiii
ABSTRACT	xiv
INTRODUCTION	1
Essential Components of an IEP	2
Professional Development and IEP Development	5
Professional Development	7
Behavior Skills Training as a PD Model	10
Quality Rubrics	12
Summary	14
Article 1: Systematic Literature Review	14
Article 2: IEP PLAAFP Development Training	15
Article 3: Practitioner Article	16
References	17
ARTICLE 1: SYSTEMATIC LITERATURE REVIEW	22
Concerns with IEP Development and Content	22
Purpose	24
Method	26
Search Procedures	26
Inclusion Criteria	28
Screener Training and Interrater Reliability	29
Screening Results and Interrater Reliability	34

Study Characteristics Coding	35
Methodological Rigor Evaluation	37
Coding Interrater Reliability	38
Data Analysis	39
Results	40
Research Question 1: Study Characteristics	42
Research Question 2: Intervention(s)	47
Research Question 3: Intervention Component(s) Effectiveness	49
Research Question 4: Addressing Concerns	51
Research Question 5: Methodological Rigor	52
Discussion	54
Implications for Practice	55
Limitations	56
Future Research	56
Conclusion	57
References	59
ARTICLE 2: IEP PLAAFP DEVELOPMENT TRAINING	64
Common PLAAFP Errors	64
Intervening on PLAAFP Quality	66
Behavior Skills Training	67
Purpose	69
Method	69
Setting	69

Participants	70
Dependent and Independent Variables	72
Materials	76
Design	76
Procedures	76
Recruitment and Participation Determination	76
Baseline	77
Intervention Phase	79
Maintenance Phase	82
Results	82
Interrater Reliability	90
Social Validity	91
Discussion	94
Implications for Practice	95
Limitations and Implications for Future Research	96
Conclusion	97
References	99
ARTICLE 3: ENHANCING PLAAFP QUALITY THROUGH THE BEHAVIOR SKILLS TRAINING (BST) MODEL	105
BST	106
Step 1: Instruction Phase- Supporting Evidence	108
Step 2: Instruction Phase- Target Skill	108
Step 3: Instruction Phase- Written Summary	111

Step 4: Modeling Phase	112
Step 5: Rehearsal Phase	114
Step 6: Feedback Phase	117
Step 7: Repeat	120
Implementation Strategies	120
Conclusion	122
References	123
DISCUSSION	125
Systematic Literature Review	127
IEP PLAAFP Development Training	128
Implications for Researchers	130
Implications for Practitioners	131
Implications for Policy Makers	132
Conclusion	133
References	135
SUPPLEMENTAL MATERIALS: Systematic Literature Review Coding Manual	139

List of Tables

CHAPTER 1: INTRODUCTION

TABLE 1: PLAAFP Components and Descriptions	4
---	---

TABLE 2: BST Coaching Steps and Descriptions	11
--	----

ARTICLE 1: SYSTEMATIC LITERATURE REVIEW

TABLE 1: Full-Text Screening Protocol	31
---------------------------------------	----

TABLE 2: IRR For Individual Methodological Rigor Indicators	39
---	----

TABLE 3: Primary Article Overview	41
-----------------------------------	----

TABLE 4: Primary Article Journal and Funding Overview	42
---	----

TABLE 5: Disability Category and Number of IEPs Reviewed	46
--	----

TABLE 6: Primary Article Independent Variable(s)	48
--	----

TABLE 7: Percentages of Indicators Met Across All Articles	53
--	----

ARTICLE 2: IEP PLAAFP DEVELOPMENT TRAINING

TABLE 1: Participant Demographics	71
-----------------------------------	----

TABLE 2: Quality Indicators for PLAAFP Statements	72
---	----

TABLE 3: Quality Indicator 2 Scoring Rubric	73
---	----

TABLE 4: De-Identification Procedures and Guidelines	78
--	----

TABLE 5: Training Video Contents and Structure	80
--	----

TABLE 6: Mean Quality Score Per Participant Per Phase	83
---	----

TABLE 7: Completed and Scored Four-Tiered Visual Analysis Protocol	87
--	----

TABLE 8: Social Validity Survey Statements	92
--	----

TABLE 9: Social Validity Survey Questions	92
---	----

TABLE 10: Social Validity Likert Scale Results	93
--	----

ARTICLE 3: ENHANCING PLAAFP QUALITY THROUGH THE BEHAVIOR SKILLS
TRAINING (BST) MODEL

TABLE 1: PLAAFP Statement Required Components	110
TABLE 2: Summary of Modeling Phase Practices	113
TABLE 3: Rehearsal Phase Components	115

List of Figures

ARTICLE 1: SYSTEMATIC LITERATURE REVIEW

FIGURE 1: PRISMA Flow Diagram	34
-------------------------------	----

ARTICLE 2: IEP PLAAFP DEVELOPMENT TRAINING

FIGURE 1: Example Feedback Form	75
---------------------------------	----

FIGURE 2: Participant Data	85
----------------------------	----

FIGURE 3: Interquartile Range	86
-------------------------------	----

ARTICLE 3: ENHANCING PLAAFP QUALITY THROUGH THE BEHAVIOR SKILLS TRAINING (BST) MODEL

FIGURE 1: BST Steps	107
---------------------	-----

FIGURE 2: Task Analysis for PLAAFP Statement Development	109
--	-----

FIGURE 3: Weekly Coaching Schedule Example	116
--	-----

FIGURE 4: BST Feedback Form Example	119
-------------------------------------	-----

DISCUSSION

FIGURE 1: Dissertation Design Summary	126
---------------------------------------	-----

Abstract

The development and implementation of Individualized Education Programs (IEPs) are crucial to a student with disabilities' access to Free and Appropriate Public Education (FAPE). However, many educators lack the necessary training and ongoing support to enhance the quality of the IEPs they develop and implement. This three-article dissertation begins with a systematic literature review to summarize the IEP professional development literature. The results of the systematic review were used to design a multiple-baseline single case study to examine an intervention package to teach special educators to write high-quality Present Levels of Academic Achievement and Functional Performance (PLAAFP) statements, which are the foundation of a student's IEP. Results showed a functional relation between the use of the Behavior Skills Training (BST) model and an increase in PLAAFP quality. The third article outlines the procedures employed in the multiple baseline study and translates results for practitioners. Finally, implications for researchers, practitioners, and policymakers are presented.

Introduction

Over 15% of public school students qualify for special education and related services, with a projected increase of one million additional students eligible for special education services by the 2025 school year (National Center for Education Statistics, 2024). Therefore, the demand for qualified educators and service providers is more pressing than ever. The Individuals with Disabilities Improvement Act (IDEA, 2004) requires public schools to support students with disabilities in accessing the necessary support and services to receive a free and appropriate public education (FAPE). Before the creation and enforcement of IDEA, *Brown v. Board of Education* (1954) marked a turning point in the nation's history. *Brown v. Board of Education* (1954) was a landmark U.S. Supreme Court decision that refuted the “separate but equal” doctrine and outlawed segregation based on race in U.S. public schools. The Court decided that “separate but equal” was inherently unequal, violating the Fourteenth Amendment, which enacted the Equal Protection Clause (*Brown v. Board of Education*, 1954). The decision set a precedent for future cases that segregation of students in public education based on demographic factors (e.g., race, sex, disability) does not provide equal access to public education, no matter the attempt to ensure that the educational facilities are “equal.” The decision was pivotal in shaping later legislation regarding the separation of students with and without disabilities. It inspired special education advocates to seek similar determinations later, as it began to shift societal views of segregation within public education.

Although there has been numerous changes to special education law and policy since *Brown v. Board of Education* (1954), including the Elementary and Secondary Education Act (ESEA, 1965), Education for All Handicapped Children Act (EHA, 1975), and several revisions of the Individuals and Disabilities Improvement Act (IDEA, 1990, 1997, 2004), the requirement

of providing an equitable education for students with disabilities still stands. With each change to special education regulations, students with disabilities have been afforded greater educational access and opportunities. The enactment of IDEA in 1990 provided schools with substantial guidance compared to prior legislation. These additional provisions included implementing procedural safeguards which outline children's educational rights and increase families' ability to advocate for their student (Individuals with Disabilities Education Act, 1990). To educate students with disabilities who are eligible for special education and related services, public schools must provide specialized instruction tailored to their needs. To fully understand the current state of special education in the US, it is essential to explore the current factors that significantly impact a special education student's ability to access FAPE. These factors include the nationwide shortage of special educators, inadequate training for educators regarding IEPs, and public schools' access to resources necessary for developing, assessing the quality of, and implementing IEPs to support students' access to FAPE.

Essential Components of an IEP

Individualized Education Programs (IEPs) comprise several key components, informed by data, team discussions, and the specific needs of each student. These components include the present level of academic achievement and functional performance (PLAAFP) levels, assessment data, measurable annual goals, accommodations and modifications, special education and related services, and detailed information regarding measuring and reporting student progress (IDEA, 2024). Each component guides the educational team on how best to support the student's academic and functional progress and assists the team in ensuring the student has equal access to a FAPE (Harmon et al., 2020).

Developing an Individualized Education Program (IEP) is a comprehensive process that begins with collecting essential assessment data. These data, which may include formal, informal, and narrative assessments, provide the educational team with the current student's level of academic and social skills to inform the creation of the IEP. Once these data are evaluated, the team can determine appropriate annual goals based on the student's educational and functional needs. These goals should adhere to the SMART criteria: Specific, Measurable, Achievable, Relevant, and Time-bound (Diehm, 2017). After establishing goals, the team must outline how progress will be continuously monitored to ensure that the chosen placement and supports remain suitable for the student's needs.

The IEP team's responsibility extends to analyzing comprehensive evaluation data to determine the most appropriate services for the student, including assessing the support needed in the general education setting. This is particularly important as research has shown that students with disabilities in general education settings often receive heavily modified work, potentially limiting their access to the same educational experiences as their non-disabled peers (Harrison et al., 2013; Joyce et al., 2018). Accommodations and modifications should be individualized and evidence-based rather than following a "one-size-fits-all" approach. Team members must review assessment data thoroughly based on student needs to determine evidence-based accommodations and modifications that truly support the appropriate inclusion of students with disabilities. This careful consideration ensures that students receive an equal educational experience and that the IEP fulfills its purpose as mandated by IDEA, providing a tailored education plan that addresses the unique needs of each student with a disability.

Once all components of the IEP have been determined, the team can then begin structuring the PLAAFP, which is the first section of the IEP, as it provides the educational team

with an overview of each section and detailed information that explains the rationale for the decisions made regarding goals, services, accommodations, and modifications (Goran et al., 2020). The PLAAFP should include a disability impact statement, student academic progress, social progress, parent and student concerns, as well as current levels of performance. Student career and professional goals should also be addressed for children at 14 or 16 years of age, depending on the state. PLAAFP statements should be strengths-based.

The courts have upheld the importance of quality IEPs. For example, Kirby v. Cabell County School District (46 IDELR 156, S.D.W. Va. 2006) found that an IEP written by the school district was invalid due to inadequate assessment. The district court judge stated, "If the IEP fails to assess the child's present levels of academic achievement and functional performance, the IEP does not comply with IDEA" (Kirby v. Cabell County Sch. Dist., 46 IDELR 156, S.D.W. Va. 2006). This ruling underscores the importance of high-quality PLAAFPs and emphasizes the value of data-driven decision-making. Assessment information is crucial in creating PLAAFPs, as all student education decisions are based on current assessment data. In addition to assessment information, PLAAFP statements should highlight the impact of the student's disability on their education, student strengths and needs, and the educational team's plan to support the student towards their goals. All of this information, though, must be based on assessment data. Table 1 provides an overview of the essential components of a high-quality PLAAFP.

Table 1

PLAAFP Components and Descriptions

PLAAFP Component	Description
Impact of Disability	Clearly explain how the student's disability affects:

	• Their involvement in the general education curriculum. • Their progress in the general education curriculum.
Assessment Data	Best practice is to include three forms of assessment data: • Formal (e.g., standardized achievement tests) • Informal (e.g., classroom activities, teacher-made assessments, observations) • Narrative (e.g., anecdotal records) For each section of assessment data, the team should include: • Context (e.g., when, where, setting) • Score (e.g., raw score, standard score, grade equivalent) • Comparison of score to non-disabled peers
Strengths and Needs	Both strengths and needs must be based on the assessment data: • Be detailed and specific • Provide clear guidance for goals, services, accommodations, and modifications • Needs should include the specific deficit • How will the need impact the student's ability to master grade-level standards? • Which academic and functional areas are impacted?
Services, Accommodations, and Modifications	Lastly, the team will outline the necessary supports: • Services • Accommodations • Modifications

Note: Content adapted from "IRIS | page 6: High-Quality PLAAFP statements. (2020).

Vanderbilt.edu. <https://iris.peabody.vanderbilt.edu/module/iep01/cresource/q3/p06/#content>

Professional Development and IEP Development

Understanding best practices for professional development is essential for ensuring that students with disabilities receive equitable access to public education. Specifically, targeted training is necessary to help special educators create high-quality and compliant IEPs that promote positive student outcomes. Most recent research emphasizes descriptive practices

regarding current IEP quality rather than interventions aimed at improving IEP quality. To effectively intervene, further research is required to outline interventions that enhance the quality of IEPs. Several descriptive studies focus on the quality of IEP components (see Burns, 2023; Goodwin et al., 2020; Hott et al., 2020; Kurth et al., 2022). For example, Hott and associates (2020) used researcher-developed rubrics based on IDEA requirements to analyze the compliance and quality of IEPs for students with math learning disabilities within rural independent school districts. The results showed that most IEPs did not include access to special education services to focus on math skill deficits identified during the evaluation process.

Burns and colleagues (2023) analyzed the quality of PLAAFP statements in correlation to student outcomes. The researchers created a PLAAFP quality rubric based on IDEA standards. In addition to analyzing PLAAFP quality, the researchers gathered student data from benchmark assessments and training information for each educator. The results revealed that PLAAFP quality was indicative of student benchmark outcomes. It was also revealed that the amount and type of professional development the educators had participated in correlated with PLAAFP quality. Special educators with access to ongoing professional development had higher quality scores, while educators who had participated in a single training session had lower PLAAFP quality scores. Goodwin and associates (2020) examined the quality of IEP goals for students with traumatic brain injuries (TBI) using the Revised IEP/IFSP Goals and Objective Rating Instrument (R-GORI). The researchers found that the average quality score was 3.47 out of 6, with 96% of goals rated as measurable; however, only 40% of goals met the criteria for generalization to different settings, materials, and interventionists.

Kurth and colleagues (2022) analyzed 112 IEPs of elementary students with complex support needs across four different educational placements. The researchers developed a quality

rating instrument to evaluate the quality of three IEP components: PLAAFPs, goals, and supplementary aids and services. The researchers determined no statistically significant differences in IEP quality among placement types. The results revealed that the IEPs consistently failed to meet the quality indicators across all placements. The mean IEP quality score was 14.4 out of 30 points (Kurth et al., 2022). PLAAFP statements had an overall mean score of 4.3 out of 10. Supplementary aids and services received an overall mean score of 3.4 out of 10, while goals were the highest-rated component with a 6.7 out of 10. More specifically, PLAAFP statements scored the lowest in data-driven practices, indicating a low use of data to develop the statements. Given the body of work suggesting IEPs do not meet procedural or substantive requirements, coupled with the teacher shortages and alternative licensure routes, the professional development needs of educators must be considered.

Professional Development

Professional development (PD) in education refers to activities and resources that help educators improve their skills, knowledge, and effectiveness (Learning Forward, 2018). This practice supports educators in gaining new skills through continuing education and training after they have entered the classroom. However, a literature review is needed to determine dosage, intensity, and PD components to further intervene in IEP quality and compliance. Donath and colleagues' (2023) synthesized the effectiveness of professional development (PD) programs in facilitating the implementation of inclusive practices within special education. The researchers assessed PD's influence on special educators' knowledge and skills and student outcomes. Analysis of the PD program specifics showed significant variation in both duration and length. PD programs lasted anywhere from 2 to 750 hours, but the median length was 20 hours (Donath et al., 2023). Regarding the duration, programs spanned from half a day to three school years,

with a median of three months. Most PD initiatives concentrated on particular subjects, primarily “special education needs.” Moreover, among the 154 programs reviewed, only 51 included additional coaching. The study presents compelling evidence that PD programs positively influence educators' knowledge, skills, and student results. Nonetheless, the observed effect sizes indicate potential for enhancements in the design and execution of PD programs.

With the current educator shortage (Peck, 2024), special education teachers are tasked with teaching and managing exponentially higher caseloads than in previous years (*January School Pulse Panel*, 2022). Determining the professional development components and package details may provide researchers and school staff with a more detailed look at the necessary components to increase quality and compliance. Doren and colleagues (2012) used a professional development model to increase the quality of transition goals. IEP data were collected to identify any correlation between goal quality, student characteristics, and teacher characteristics. The researchers found that the professional development model used increased the quality of transition goals, and the years of experience and classroom type of the teacher played a significant role in the quality of the goals.

Most recently, several studies have attempted to intervene and increase the quality and compliance of IEPs (Müller et al., 2022; Rakap, 2023; Rakap & Balikci, 2024). Müller (2023) used a data-driven social-emotional learning (SEL) IEP goal builder to intervene and increase the quality of SEL goals. Results showed that the IvySCIP goal builder increased the quality of SEL goals. The goals written using the goal builder included more required goal components than before using the builder. This study focused on the SEL goals, while other researchers attempted to intervene in IEP goals as a whole (Rakap, 2023; Rakap & Balikci, 2024; Shriner et al., 2012). Rakap (2023) and Rakap & Balikci (2024) used the artificial intelligence (AI) program ChatGPT

to increase the quality and compliance of IEP goals. More specifically, Rakap (2023) used ChatGPT to improve the quality of IEP goals for novice special education teachers. The researchers provided participants with short training on goal writing and a knowledge test before using the AI system to create goals. Each participant was required to reach a certain mastery level, determined by the knowledge test, before beginning the study.

Currently, IEP studies require extensive time from both the intervention agent and the participants. Shriner and colleagues (2012) implemented a web-based decision-making tool to enhance the quality of IEP goals created by special educators. The participants were required to attend a three-hour on-site training session on how to use the web-based tool. Subsequently, they needed to submit finalized IEPs to the research team, in addition to adhering to their normal school and district procedures for finalizing and submitting IEPs. Participants also had access to ongoing support through email and phone communication with the professional development interventionist. Overall, the authors found that educators who utilized most of the online tools and resources significantly improved the quality of IEPs. Although the results indicated a positive association between the use of online resources and an increase in IEP quality, the authors noted that they did not control the progression of IEP outcomes, suggesting that intervening or mediating variables may have influenced the effects of online resources on IEP quality.

Lowman (2015) placed participants in three intervention groups, each requiring different levels of commitment. The first group received training content via a website, the second was required to attend a full-day workshop that included peer coaching as a form of professional learning, and the third group attended the same full-day workshop as the second group but had the additional requirement of follow-up peer coaching through an online platform. The

researchers found little to no difference between the second and third groups, indicating that the addition of continued peer coaching did not enhance the quality of IEP goals. The results emphasized the importance of providing ongoing and easily accessible support for IEP development through web-based resources (Lowman, 2015). Furthermore, the findings stressed the need for professional development programs to highlight the connection between IEP quality and instructional planning and implementation, and the necessity to consider educator experience levels when designing PD programs.

Conversely, Rakap (2023) provided the participants with written guidance on using ChatGPT to create SMART goals, requiring minimal additional effort and time commitment from the participants. Although the intervention demanded little to no extra workload for the participants, they were still required to submit IEP goals to the researchers, in addition to the regular school and district submission requirements. The researchers also found that educators with prior training in IEP development and goal writing positively influenced the quality of IEP goals created by early-career special educators. While the designed PD program required minimal additional work for the participants, it did raise ethical concerns. When considering the use of AI tools in PD programs, especially for IEP development, one should reflect on ethical implications such as privacy, data security, and potential biases (Rakap, 2023)

Behavior Skills Training as a PD Model

The Behavior Skills Training (BST) model was created to provide a systematic approach to coaching and training based on empirical evidence (DiGennaro Reed et al., 2018; Hogan et al., 2014). The BST coaching model involves four systematic steps: instruction, modeling, rehearsal, and feedback (Kirkpatrick et al., 2019). The instruction step is conducted first to provide participants with the necessary information about the topic of interest. This step provides

examples, non-examples, and detailed descriptions and definitions. Trainers should be well-versed in the content and provide participants with verbal and written instructions. The modeling step can be completed independently or during the instruction phase. This step provides the participants with a visual example of how the skill being taught should be done. Modeling can be done in person or in vivo. In-vivo modeling can allow participants to review the video later as a continued reference to support implementation fidelity. The third step in the BST model is rehearsal, where the participant practices the skill. DiGennaro Reed and associates (2018) recommend integrating constructive feedback within the rehearsal steps to increase effectiveness. Therefore, the rehearsal and feedback steps are typically combined when the content is initially taught to participants. Delayed feedback can be delivered after the initial rehearsal and feedback sessions to support implementation fidelity and correct any drift. All feedback should be constructive and focus on strengths and improvement areas. As previously mentioned, verbal and instantaneous feedback should be used in conjunction with the first rehearsal phase, while written and delayed feedback can be used in later rehearsal and feedback phases. Table 2 provides a breakdown of the BST coaching model.

Table 2

BST Coaching Steps and Descriptions

BST Model Steps	Description
Instruction	Participants, or trainees, receive brief training on the skills of interest. This initial step is typically lecture-based but can be adapted to meet the trainees' specific needs.
Modeling	Next, the trainer models the skills for the trainees. This can occur during the instruction step or in separate training sessions. It is important that the modeling session takes place within a reasonable timeframe after the instruction has been delivered.

Rehearsal	Following this, trainees practice the skill in either a controlled or natural environment, depending on their individual needs, the availability of the environment, or the specific skill being targeted. During this practice, the trainer provides on-the-spot coaching to support the trainee.
Feedback	After the practice session, the trainer offers constructive feedback, highlighting both strengths and areas for improvement. Based on the trainee's performance during the practice phase, the trainer may choose to revisit any of the previous three steps as needed.

Note. Adapted from “Guidelines for Using Behavioral Skills Training to Provide Teacher Support” DiGennaro Reed, F. D., Blackman, A. L., Erath, T. G., Brand, D., & Novak, M. D., 2018, *TEACHING Exceptional Children*, 50(6), 373–380 (<https://doi.org/10.1177/0040059918777241>).

Quality Rubrics

IEP studies have employed various rubrics to assess whether the standards necessary for compliance and effectiveness are evident within the IEP (Burns et al., 2023; Doren et al., 2012; Lowman, 2016; Rakap, 2023; Ruble & McGrew, 2013). It is common practice for researchers to create a quality rubric specific to the IEP area that is being analyzed within the study (see Burns et al., 2023; Kurth et al., 2022; Lowman, 2016; Ruble & McGrew, 2013). Although these quality rubrics are often created by special education professionals and based on IDEA and state requirements, many still need to be tested to determine if the tool is valid before use within the study. Validity, specifically content validity, refers to how well a rubric measures what it is intended to measure. There are several ways to determine content validity, such as the Table of Specifications method, originally based on Lawshe's (1975) work and recently updated by Newman and associates in 2013. With this method, researchers and educational professionals evaluate the extent to which each indicator aligns with major content themes that foster high-

quality IEP development. Once the indicators are assessed, the results are compared to Lawshe's Content Validity Ratio to determine the range of validity (Lawshe, 1975). In addition to validity, quality rubrics should also be evaluated for reliability, which refers to the consistency of scores when using the rubric (Ruble et al., 2010). Reliability can be determined through interclass correlation coefficients (ICC); the higher the score, the more reliable the measure is (Ruble et al., 2010).

Using a validated quality tool is essential to the study's results and the reader's ability to generalize them to everyday practice. With a validated IEP quality tool, one can be assured that the quality percentages are accurate and that the expectations outlined in federal and state mandates are being met. Several research studies on IEP quality have focused on transition goals, utilizing validated quality tools disseminated by federal and state organizations (Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015). The tools commonly used are the Transition Requirements Checklist created by Storms, O'Leary, and Williams in 2000. It was later adapted in 2001 by O'Leary, Lehman, and Doty. Although this quality tool has been validated through empirical research, it was created and adapted prior to the 2004 reauthorization of IDEA, which introduced substantial changes to the language and requirements surrounding transition goals and services.

The Revised IEP/IFSP Goals and Objective Rating Instrument (R-GORI; Notari-Syversion & Schuster, 1995) and the IEP Quality Indicator Scale for Goals/Objectives (IQIS-Goals/Objectives; Yell et al., 2008) are empirically validated IEP quality rubrics. Rakap (2023) and Rakap & Balikci (2024) used the R-GORI to determine the quality of IEP goals after the participants, special educators, used ChatGPT, an artificial intelligence (AI) program, to assist in the creation of IEP goals. Although both studies showed promising results, the R-GORI quality

rubric was developed in 1995, prior to the 2004 reauthorization of IDEA. Shriner and associates (2012) utilized the IQUIS-Goals/Objectives to assess the quality and compliance of IEP goals created by participants and special educators using a web-based decision-making tutorial.

Summary

The introduction provides an overview of IEP procedural, substantive, and instructional issues, indicating that there is a critical need for comprehensive professional learning to enhance the quality and compliance of IEPs. Despite the legal requirements and importance of IEPs in ensuring educational opportunity, educators benefit from professional development to develop and implement effective plans (Finn & Kohler, 2009; Kurth, 2022). Common challenges include a need for more awareness about IEP components (Burns, 2023), insufficient resources and support (Yam, 2023), and difficulty writing measurable goals aligned with standards (Ruble & McGrew, 2013). Additionally, even when quality IEPs are created, there are a myriad of implementation challenges. This gap between IEP development and implementation can significantly impact students' academic progress and access to needed accommodations. Targeted professional learning is essential to build educators' capacity to collaboratively create meaningful, standards-aligned IEPs and successfully implement them to support positive outcomes for students with disabilities. This three-article dissertation comprises a systematic literature review of studies that have intervened on IEP quality, a single-case research design that intervenes on the quality of PLAAFP statements, and a practitioner article that outlines the coaching methods employed in the second study.

Article 1: Systematic Literature Review

The first study is a systematic literature review (Hott et al., 2021, pp. 89–111). This review aims to gather more information on the research-identified methods for intervening and

enhancing IEP quality. Systematic reviews of IEP studies have focused on IEPs for students with learning disabilities (McKenna et al., 2023) and methods to increase student participation in IEP meetings (Sanderson et al., 2020, 2022). McKenna and colleagues synthesized IEP quality research specific to students with specific learning disabilities (SLD). Sanderson and associates (2022) reviewed single-case research aimed at increasing student participation in IEP meetings, along with a review of all methods utilized to boost student participation (2020). Currently, no systematic reviews encompass all disability categories outlined by IDEA and focus on the quality of one or more IEP components. The remaining disability categories are autism, deaf-blindness, deafness, emotional disturbance, hearing impairment, intellectual disability, multiple disabilities, orthopedic impairment, other health impairment, speech or language impairment, traumatic brain injury, and visual impairment. The first article synthesizes the literature that either analyzes or intervenes in the quality of one or more IEP components. The results guided the researcher in determining the evidence-based practices to be used in the second study. Information such as duration, length, training components, and effectiveness was gathered to assist the researcher in designing and completing the second study.

Article 2: IEP PLAAFP Development Training

The second study used the results from the systematic review to inform the development of an intervention aimed at enhancing the quality of IEP PLAAFPs for elementary-aged students. The details of the interventions and the effectiveness reported in the studies identified through the systematic review informed the researcher's creation and implementation of the intervention. The study employed an independent delayed multiple baselines across participants design (Ledford & Gast, 2018) to investigate whether there is a functional relationship between using the BST coaching model (DiGennaro Reed et al., 2018) and improving the quality of PLAAFP

statements written by special educators in high-needs elementary schools. By leveraging evidence-based practices and the Behavior Skills Training (BST) model, this intervention aimed to enhance educators' understanding of the Individuals with Disabilities Education Act (IDEA) requirements and improve their ability to make data-driven and comprehensive PLAAFP statements. Through instruction, modeling, rehearsal, and feedback, participants developed the capacity to produce PLAAFP statements that comply with legal standards and promote meaningful educational outcomes and inclusion for all students.

Article 3: Practitioner Article

The third article incorporated the results from the systematic literature review and the professional development study to translate findings in a format accessible to practitioners. Specifically, the researcher provides an outline and detailed guidance for practitioners in developing coaching models to support teachers in developing high-quality PLAAFPs.

References

- Brown v. Board of Education, 347 U.S. 483 (1954)
- Diehm, E. (2017). Writing measurable and academically relevant IEP goals with 80% accuracy over three consecutive trials. *Perspectives of the ASHA Special Interest Groups*, 2(16), 34–44. <https://doi.org/10.1044/persp2.sig16.34>
- DiGennaro Reed, F. D., Blackman, A. L., Erath, T. G., Brand, D., & Novak, M. D. (2018). Guidelines for using behavioral skills training to provide teacher support. *TEACHING Exceptional Children*, 50(6), 373–380. <https://doi.org/10.1177/0040059918777241>
- Donath, J. L., Lüke, T., Graf, E., Tran, U. S., & Götz, T. (2023). Does Professional Development Effectively Support the Implementation of Inclusive Education? A Meta-Analysis. *Educational Psychology Review*, 35(1). <https://doi.org/10.1007/s10648-023-09752-2>
- Education for All Handicapped Children Act, Pub. L. No. 94-142, 89 Stat. 773 (1975)
- Elementary and Secondary Education Act, 20 U.S.C. § 6301 et seq. (1965)
- Finn, J. E., & Kohler, P. D. (2009). A Compliance Evaluation of the Transition Outcomes Project. *Career Development for Exceptional Individuals*, 32(1), 17–29. <https://doi.org/10.1177/0885728808315332>
- Goran, L., Harkins Monaco, E. A., Yell, M. L., Shriner, J., & Bateman, D. (2020). Pursuing academic and functional advancement: Goals, services, and measuring progress. *TEACHING Exceptional Children*, 52(5), 333–343. <https://doi.org/10.1177/0040059920919924>
- Harmon, S., Street, M., Bateman, D., & Yell, M. L. (2020). Developing present levels of academic achievement and functional performance statements for iep. *TEACHING Exceptional Children*, 52(5), 320–332. <https://doi.org/10.1177/0040059920914260>

- Harrison, J. R., Bunford, N., Evans, S. W., & Owens, J. S. (2013). Educational accommodations for students with behavioral challenges. *Review of Educational Research*, 83(4), 551–597. <https://doi.org/10.3102/0034654313497517>
- Hott, B. L., Jones, B. A., Randolph, K. M., Kuntz, E., McKenna, J. W., & Brigham, F. J. (2020). Lessons learned from a descriptive review of rural individualized education programs. *The Journal of Special Education*, 55(3), 002246692097267. <https://doi.org/10.1177/0022466920972670>
- Hott, B. L., Brigham, F. J., & Peltier, C. (2021). *Research methods in special education* (pp. 89–111). Slack Incorporated.
- Individuals with Disabilities Education Act Amendments of 1997, Pub. L. No. 105-17, 111 Stat. 37 (1997)
- Individuals with Disabilities Education Act, Pub. L. No. 101-476, 104 Stat. 1103 (1990)
- Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446, 118 Stat. 2647 (2004)
- Joyce, J., Harrison, J. R., & Gitomer, D. H. (2018). Modifications and accommodations: a preliminary investigation into changes in classroom artifact quality. *International Journal of Inclusive Education*, 24(2), 1–21. <https://doi.org/10.1080/13603116.2018.1453876>
- Kirby v. Cabell County School District, 46 IDELR 156 (S.D.W. Va. 2006).
- Kirkpatrick, M., Akers, J., & Rivera, G. (2019). Use of behavioral skills training with teachers: A systematic review. *Journal of Behavioral Education*, 28(3), 344–361. <https://doi.org/10.1007/s10864-019-09322-z>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.

Learning Forward. (2018, July 25). *Professional Learning Definition* | *Learning Forward*.

Learning Forward - the Professional Learning Association.

<https://learningforward.org/about/professional-learning-definition/?t>

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

Lowman, J. J. (2015). A comparison of three professional development mechanisms for improving the quality of Standards-Based IEP objectives. *Communication Disorders Quarterly*, 37(4), 211–224. <https://doi.org/10.1177/1525740115595338>

McGuire, S. N., & Meadan, H. (2022). General educators' perceptions of social inclusion of elementary students with emotional and behavioral disorders. *Behavioral Disorders*, 48(1), <https://doi.org/10.1177/01987429221079047>

McKenna, J. W., Solis, M., Garwood, J., & Parenti, M. (2023). Characteristics of individualized education programs for students with learning disabilities: A systematic review. *Learning Disability Quarterly*, 1. <https://doi.org/10.1177/07319487231182697>

National Center for Education Statistics. (2024). Students With Disabilities. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved 2025, from <https://nces.ed.gov/programs/coe/indicator/cgg>.

Newman, I., Lim, J., & Pineda, F. (2013). Content validity using a mixed methods approach: Its application and development are done through the use of a table of specifications methodology. *Journal of Mixed Methods Research*, 7(3), 243-260. doi: 10.1177/1558689813476922

Peck, D. (2024). *The Ultimate List of Teacher Shortages by State in 2024* | *Devlin Peck*. Devlinpeck.com. <https://www.devlinpeck.com/content/teacher-shortage-by-state?t>

- Rakap, S. (2023). Chatting with GPT: Enhancing individualized education program goal development for novice special education teachers. *Journal of Special Education Technology*. <https://doi.org/10.1177/01626434231211295>
- Rakap, S., & Balikci, S. (2024). Enhancing IEP goal development for preschoolers with autism: A preliminary study on ChatGPT integration. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-024-06343-0>
- Ruble, L. A., McGrew, J., Dalrymple, N., & Jung, L. A. (2010). Examining the quality of IEPs for young children with autism. *Journal of Autism and Developmental Disorders*, 40(12), 1459–1470. <https://doi.org/10.1007/s10803-010-1003-1>
- Ruble, L., & McGrew, J. H. (2013). Teacher and child predictors of achieving IEP goals of children with autism. *Journal of Autism and Developmental Disorders*, 43(12), 2748–2763. <https://doi.org/10.1007/s10803-013-1884-x>
- Sanderson, K. A., & Goldman, S. E. (2020). A systematic review and meta-analysis of interventions used to increase adolescent IEP meeting participation. *Career Development and Transition for Exceptional Individuals*, 43(3), 157–168. <https://doi.org/10.1177/2165143420922552>
- Sanderson, K. A., Goldman, S. E., & Rojas, A. (2022). A review and analysis of single-case research examining adolescent participation in IEP meetings. *The Journal of Special Education*, 56(2), 97–109. <https://doi.org/10.1177/00224669211047232>
- Shriner, J. G., Carty, S. J., Rose, C. A., Shogren, K. A., Kim, M., & Trach, J. S. (2012). Effects of using a web-based individualized education program decision-making tutorial. *the Journal of Special Education/the Journal of Special Education*, 47(3), 175–185. <https://doi.org/10.1177/0022466912453940>

Yam. (2023, August 14). *Addressing the Significant Shortage of Special Education Teachers in California*. The Arc of California. <https://thearcca.org/addressing-the-significant-shortage-of-special-education-teachers-in-california/>

Article 1

Systematic Literature Review

From the 2013-2014 school year to the 2022-2023 school year, the number of students served through special education programs increased by 15%, rising from 6.3 to 7.5 million children (National Center for Education Statistics, 2024). Kindergarten through 12th-grade students eligible for special education and related services must have an Individualized Education Program (IEP). An IEP is a legally binding document that provides school staff with a detailed, data-driven plan to ensure that students with disabilities receive a free and appropriate public education (FAPE; Individuals with Disabilities Education Act [IDEA], 2024). IEPs encompass several components that guide school staff in delivering services to effectively meet the needs of students with disabilities attending public schools. These components include (1) present level of academic and functional performance (PLAAFP), (2) measurable and attainable annual goals, (3) special education and related services, (4) accommodations, and modifications based on accurate and recent data collected through formal and informal assessments.

Concerns with IEP Development and Content

Recent research (e.g., Srivastava, 2024) emphasizes the importance of high-quality IEPs concerning student outcomes. Srivastava (2024) employed a mixed-methods approach to assess the impact of IEP quality on the academic performance of students with disabilities. The results indicated that (1) individualized lesson planning and instructional strategies, (2) family engagement, and (3) resource and support provision significantly contribute to improved student outcomes. Furthermore, the researchers concluded that two factors were essential for ensuring equitable access and promoting student success: active parental involvement in the IEP process and ongoing professional development for special educators. Recent studies (e.g., Hott et al.,

2020; Hott et al., 2021; Klang et al., 2016) have also identified shortcomings in IEPs. Klang and colleagues (2016) examined the IEPs of 43 students with complex communication needs. The researchers reviewed the communication and language goals of each IEP and compared them to the requirements set by the International Classification of Functioning, Disability, and Health for Youth (ICF-CY). The results indicated that most goals reviewed included elements from multiple ICF-CY components. A relatively small number of goals focused on peer interaction or classroom participation, essential aspects of the ICF-CY. Klang and colleagues (2016) highlighted the significance of conversational participation per ICF-CY standards, as it is an essential skill for overall independence and functioning. Although individual skills are important, students also need to develop their conversational participation skills. Without actively participating in classroom activities and interacting with peers, students are less likely to develop the skills necessary for meaningful access to general education and to acquire the skills required for a successful transition to higher grades. These findings highlight the limitations of IEP goals, including their unmeasurable and unattainable nature.

More recently, Hott and colleagues (2020, 2021) conducted two descriptive reviews of IEPs for students in rural public school settings. Hott and associates (2020) reviewed 89 IEPs for students with Mathematics Learning Disabilities (MLD) across 15 rural districts. The researchers developed coding protocols to evaluate the quality of PLAAFP statements, goals, accommodations, and services. The findings revealed that very few IEPs included goals that addressed grade-level math content, with the majority focusing on basic math calculation skills. Additionally, many accommodations lacked specificity in their implementation (e.g., extended time, reduced assignments, speech-to-text). This raises concerns since most students receive math instruction in the general education setting, where accommodations are to be implemented.

The second descriptive review examined the quality of PLAAFP statements, IEP goals, PLAAFP-goal alignment, and progress monitoring practices (Hott et al., 2021). The researchers analyzed the contents of 133 IEPs from 7 rural school districts. The quality coding protocol was based on common IEP errors identified by Yell and associates (2013). The results indicated that less than 6% of the analyzed IEPs featured complete PLAAFP statements, only 34% contained a full statement of need supporting the selected special education services, and 52% of the IEPs did not present updated assessment data but instead relied solely on teacher observations for determinations (Hott et al., 2021). The deficiency in IEP quality poses a risk to district legal compliance and, perhaps more importantly, the quality of student instruction. For instance, misaligned and incomplete PLAAFP statements may result in unsuitable and unnecessary service and placement decisions (LaSalle et al., 2013). Furthermore, unmeasurable, unclear, and misaligned annual goals could create complications with IEP development and hinder the school team's ability to monitor and adjust goals and services (Hott et al., 2020; Hott et al., 2021)

Purpose

There is a documented need for IEP development intervention through numerous descriptive and document analyses that revealed low-quality and low-compliance IEP components (Müller et al., 2022; Rakap, 2023; Rakap & Balikci, 2024). Although previous research has predominantly explored transition plans and goals (Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015; Lowman, 2015; Rakap, 2023), more actionable research is needed to improve IEP quality and compliance (Goodwin et al., 2020; Kurth & Mastergeorge, 2009; Ruble et al., 2010). Prior systematic reviews have focused on disability categories, such as specific learning disabilities (SLD; McKenna et al., 2023) and student participation in IEP meetings (Sanderson et al., 2020; Sanderson et al., 2022). Although these studies synthesize the

literature for each topic, no current systematic reviews encompass all disability categories and target interventions to improve IEP quality. Such a synthesis is necessary to provide a comprehensive overview of the interventions used to enhance IEP quality. The generalizable findings could equip researchers, practitioners, and policymakers with practical tools for improving IEP quality, reducing cost, and emphasizing time efficiency. Additionally, a systematic review that assesses methodological rigor is essential to inform readers whether the interventions implemented in previous research meet the standards to be considered evidence-based.

We aim to extend IEP research syntheses and contribute to the literature base in several ways. First, all special education placement categories are included in the synthesis. Second, this investigation aims to replicate and expand a prior systematic review conducted by McKenna et al. (2023) of peer-reviewed studies intervening on the quality of IEPs for students receiving special education services within a K-12 public education system, while also analyzing the methodological rigor of the studies to reveal the evidence-based interventions. Additionally, we intend to use the systematic review results to create an evidence-based professional learning opportunity for special education teachers and develop a practitioner article to communicate outcomes in a manner that is accessible to practitioners and policymakers. The research questions for the systematic review are as follows:

1. What are the trends in participant demographics, geographic locations, authors, funding sources, and journals across studies focused on improving IEP quality?
2. What interventions have been implemented to enhance IEP quality since the 2004 reauthorization of IDEA?
3. What common intervention components or strategies demonstrate the strongest evidence for improving IEP quality?

4. To what extent do the identified interventions address the specific areas of IEP quality concerns highlighted in recent literature?
5. What is the methodological rigor of the studies examining IEP quality interventions as evaluated by the Council for Exceptional Children (CEC) quality indicators (2014)?

Method

A systematic review provides researchers and readers with a detailed view of the literature surrounding a topic (Munn et al., 2018). In addition, systematic reviews can analyze the quality, commonalities, and differences across the literature (McKenna & Peltier, 2021b). We conducted a systematic review to consolidate the evidence supporting the intervention in IEP quality and compliance. The review assesses the methodological rigor of studies examining IEP quality and compliance. It provides researchers and readers with a compiled report on the interventions used to support quality or compliance within individualized education programs in public schools across the United States.

Search Procedures

Researchers followed the PRISMA multi-step process for new systematic reviews: searching databases and registers to identify the articles, identifying articles, screening the articles to determine which should be excluded based on the inclusion and exclusion criteria, and then reviewing the included articles (Page et al., 2021). First, an electronic search through the information platform EBSCO of the following databases was conducted: ERIC, Academic Search Premier, Education Full Text (H.W. Wilson), and Psychology and Behavioral Sciences Collection. The search was conducted in February 2025 to include articles published between 2004-2024 using the following Boolean string to search titles and abstracts: (“individualized education program” OR “IEP” OR “accommodations for students with disabilities” OR

“modifications for students with disabilities,” OR “free appropriate public education”). The Boolean string was designed to give researchers a comprehensive overview of the literature on IEPs. The search was truncated to 2004 due to the last reauthorization of IDEA, completed in 2004, which increased the requirements of IEPs and introduced language on evidence-based practices.

The initial search procedures were replicated and expanded upon from a prior systematic review conducted by McKenna and colleagues (2023) and expanded by including all disability categories under IDEA. Limiters placed on the search included (a) the article was written in English, (b) the article was peer-reviewed, and (c) the article was published in an academic journal. The limiters were chosen based on the review's purpose of locating articles within the research base that pertain to IEP quality through analysis of IEP components and determining the evidence-based practices used to increase IEP quality. Researchers screened the titles and abstracts based on the initial inclusion criteria, with the addition of the double screening being conducted by a research assistant, which was used to determine the extent of the literature base. The researcher is a doctoral student who has conducted multiple systematic reviews and undergone extensive training through coursework. The research assistant was trained in the review procedures and had experienced additional training through educational research coursework.

A first-author search of the articles chosen for the review covering the same timeframe was also completed. Next, the authors conducted a forward search of each primary article chosen for final coding. A forward search identifies articles that have cited an original article after publication (TMC Library, 2020). The forward search process included identifying primary articles, utilizing a citation database to find recent articles that have cited those primary articles,

and then reviewing the cited articles to determine their inclusion or exclusion (Hott et al., 2021, pp. 99–100). The citation databases used included Google Scholar and Web of Science. This process was performed for each author of the primary articles, and a first-author search and forward search for each primary article. Lastly, an ancestral search was run to identify additional studies cited in each primary article. An ancestral search involves reviewing the references cited by the primary articles to identify any additional studies that might meet the inclusion criteria (Hott et al., 2021, pp. 99–100).

Inclusion Criteria

The inclusion criteria were ordered by priority to allow for more narrow screening results. The initial inclusion criteria are as follows: (a) the study was conducted in the United States, (b) in a PK-12 public education setting, and (c) IEP components were analyzed and compared to baseline data, policy, procedures, and quality standards. In addition, articles were only selected for inclusion if they adhered to a research method, meaning no opinion, legal briefs, or practitioner pieces were included. The rationale for the inclusion criteria concerning research methods was that only articles analyzing some form of numerical data were considered. While practitioner articles provide valuable practical solutions to educational needs, their information depends on existing research articles rather than new or novel findings. As a result, the research supporting the practitioner piece may have already been included in the articles from the initial search, potentially duplicating information. These inclusion criteria, in addition to the previous limiters, were chosen to determine the research base on IEP quality. This allowed the research team to synthesize the information gathered to determine the current state of IEP quality research.

Following the initial search to determine the research base, additional exclusion criteria were added: (a) correlational, (b) descriptive, (c) qualitative, (d) case study, (e) case report, (f) cross-sectional design, (g) ecological design, (h) case law, (i) practitioner articles, (j) systematic review, (k) scoping review, (l) meta-analysis, and (m) surveys. The rationale for the additional exclusions was to ensure that the final articles selected implemented an intervention, as the purpose of the systematic literature review was to gain a deeper understanding of what the literature has defined as evidence-based. Furthermore, for an intervention to be regarded as evidence-based, substantial supportive experimental research must positively affect student outcomes (Talbot et al., 2017).

Screening Training and Interrater Reliability

Before training the research assistant, the first author screened the first 50 articles from the initial search results. Then, the research assistant was trained in the screening criteria and process in person. The training was approximately one hour in duration and was based on the needs of the research assistant. During the training, the research assistant was provided with detailed inclusion/exclusion criteria that contained definitions, examples, and non-examples. Once the screening criteria and process training were concluded, the research assistant was asked to screen 50 titles and abstracts, the same 50 that the first author initially screened, to determine the reliability of the screening criteria and procedures. The research assistant screened the provided titles and abstracts, and the first author calculated inter-rater reliability (IRR). Interrater reliability (IRR) was determined using total agreement and unscored agreement percentages (Vollmer et al., 2008). Total agreement IRR was chosen based on its ease of calculation and to provide the researchers with a broad consistency check. The IRR for the first 50 titles and abstract screening was 100%. The research assistant was then provided with the remaining titles

and abstracts to meet the 30% double-screening. IRR for all double-screening was 100%; therefore, no consensus meetings were needed.

Once the initial screening was complete, the articles selected to meet the title and abstract inclusion criteria underwent a full-text review. The researcher and research assistant conducted a full-text review of all included articles. For each article, a screening sheet that included the inclusion/exclusion criteria and examples for each criterion was completed, and the criteria met and not met were indicated. The article had to satisfy all inclusion criteria to be considered a primary article. The same IRR methods were used during the screener training. Table 1 provides a screening protocol that will be used for full-text review.

Table 1*Full-Text Screening Protocol*

Inclusion	Exclusion
1. Published in English 2. Study conducted in the United States 3. PK-12 Public Education Setting (4 years to 22 years old) 4. Study Method: Single Case Research Design: Researchers administer the intervention and collect data for an individual case (such as a child or family). Each case serves as its own control, so there is no need for a comparison group in the study design. The outcome is measured repeatedly throughout the study. Group design studies compare outcomes between units—such as students, teachers or schools—in two or more groups. Units in at least one group receive the intervention—the programs, policies, products or practices. At least one other group is a comparison group, which did not implement the intervention. A mixed methods research design is a procedure for collecting, analyzing, and “mixing” both quantitative and qualitative research and methods in a single study to understand a research problem. Only quantitative data will be reviewed and included in the systematic review. Intervention Studies:	1. Published in a language that is not English 2. Study conducted outside of the United States 3. Private Education Setting (Charter School, Private School, Clinic, Day School, Hospital) 4. Study Method: Qualitative research design typically involves gathering data through methods such as interviews, observations, focus groups, and analysis of documents or artifacts. These methods allow researchers to collect detailed, descriptive information about participants' perspectives, experiences, and contexts. A case report refers to the description of a patient with an unusual disease or with simultaneous occurrence of more than one condition. A case series is similar, except that it is an aggregation of multiple (often only a few) similar cases. Many case reports and case series are anecdotal and of limited value. Studies with a cross-sectional study design involve the collection of information on the presence or level of one or more variables of interest (health-related characteristic), whether exposure (e.g., a risk factor) or outcome (e.g., a disease) as they exist in a defined population at one particular time. If these data are analyzed only to determine the distribution of one or more variables, these are “descriptive.”

-
- Independent Variable: is an intervention to increase quality and/or compliance
 - Dependent Variable: is the quality or compliance of an IEP component.

Ecological (also sometimes called as correlational) study design involves looking for association between an exposure and an outcome across populations rather than in individuals.

Case law, also used interchangeably with common law, refers to the collection of precedents and authority set by previous judicial decisions on a particular issue or topic.

Practitioner Articles

These are frequently based on real-life experiences or observations—are primarily written by a professional in a given field to support colleagues in a variety of ways including advising, establishing best practices, and noting trends.

Systematic reviews, as the name implies, typically involve a detailed and comprehensive plan and search strategy derived a priori, with the goal of reducing bias by identifying, appraising, and synthesizing all relevant studies on a particular topic.

Scoping reviews aim to answer questions that are often exploratory in nature, such as descriptions of professional identity and clinical reasoning, and require the use of synthesis methods. Scoping reviews include different types of literature and cast a wide net for mapping the literature.

Meta-analysis is the statistical combination of the results of multiple studies addressing a similar research question. An important part of this method involves computing an effect size across all of the

studies; this involves extracting effect sizes and variance measures from various studies.

A correlational study is a type of research design that looks at the relationships between two or more variables. Correlational studies are non-experimental, meaning the experimenter does not manipulate or control any variables. A correlation refers to a relationship between two variables.

Survey research is defined as "the collection of information from a sample of individuals through their responses to questions" (Check & Schutt, 2012, p. 160). This type of research allows for a variety of methods to recruit participants, collect data, and utilize various methods of instrumentation

IEP components are discussed and analyzed, when compared to baseline data, policy and procedures, or other quality standards.

IEP components are discussed, but not analyzed or compared to baseline data, policy and procedures, or other quality standards.

Screening Results and Interrater Reliability

The initial search produced 8,778 articles, with 3,701 remaining after removing duplicates. The journal search concluded that no additional articles had been found. After title and abstract screening, 38 articles were chosen for full-text review. All 38 articles that met the initial inclusion criteria underwent a full-text review and were independently read by the first author and research assistant, with notes taken on the inclusion and exclusion criteria met. After the full-text review, nine articles (23.7%) were determined to meet all inclusion criteria, becoming primary studies for the review. Articles were commonly excluded for lacking an analysis of IEP components for the participants and not manipulating the independent variable to increase IEP quality. Figure 1 illustrates the article determination process.

After double-screening, interrater reliability for the total agreement was 99.85% and 99.95% for the unscored agreement. The first author and research assistant reviewed the individual screening results, discussed any disagreements, and agreed to a consensus. Concluding the full-text review, all primary articles underwent a full-text review by the first author and research assistant; interrater reliability for the total agreement was initially 97% and 100% after an agreement meeting, leading to the nine primary articles in the systematic review. Additionally, a forward search of all authors of the primary articles and a backward search of the primary articles concluded no additional articles meeting the inclusion criteria were found.

Figure 1

PRISMA Flow Diagram



Study Characteristics Coding

The following study characteristics were coded: (1) year of publication, (2) study authors, (3) funding statements, (4) journal, (5) research design type, (6) independent variable, (7) dependent variable, (8) intervention package details, (9) participant demographics, and (10)

locale information. The journal and all authors were coded to determine the overlap between journals and research teams that conduct research and publish work on IEP quality. Design type, independent variable, dependent variable, and intervention package were coded to provide the author with details on evidence-based practices to support IEP quality and guide the design of the single-case research design in the second article. This information also helped determine any overlap in interventions used within the research base to enhance IEP quality. Information was collected, such as the number of IEPs analyzed, the IEP quality rubric used, and the IOA for IEP quality coding. Another characteristic component collected from the primary studies was funding statements, which were used to determine overlaps and saturation within funding types (e.g., federal, state, private, and foundation).

Lastly, participant demographics and locale information were coded to assist the authors in determining saturation and gaps within the research base regarding the generalizability of the interventions being used. Participant demographics were broken into four categories: student treatment group, student control group, teacher treatment group, and teacher control group. Demographics, including primary disability, ethnicity, age, years of experience, highest level of education, and grade level, were coded for all four groups as applicable. Locale information, such as geographical location, location type (e.g., rural, suburban, and urban), and locale descriptors (e.g., large county, mid-size city), was coded. Each section was coded as either a (1) yes, the component was included, or (0) the component was not included; the same coding procedures were completed for locale information and quality indicators. No partial points were coded, as partial inclusion will be documented through the additional information gathered during the coding process. This information was pulled from each primary article for later descriptive analysis.

Each coder entered information on all study characteristics into a Qualtrics survey. The first author created the Qualtrics survey, which was then reviewed by expert methodologists in special education research. Adjustments to assist in flow were made based on the expert methodologist's review. The first author field-tested the survey, adapted it, and then reviewed it with an expert methodologist for a final time. The final survey, containing the study characteristics and additional coding measures, was used to collect data for all primary studies. The full coding manual is provided in the supplemental materials.

Methodological Rigor Evaluation

The authors evaluated each primary article for methodological rigor according to the group comparison design standards from the Council for Exceptional Children's Standards for Evidence-Based Practices in Special Education (2014). The indicators of group comparison study rigor can assess randomized controlled trials (RCTs) and non-random assignments, including quasi-experimental and regression discontinuity designs. The CEC methodological rigor indicators were selected due to the specifics of the systematic review and the constraints placed on the original search. Moreover, several compelling reasons exist for choosing CEC standards over alternatives like the What Works Clearinghouse methodological rigor standards. First, the CEC standards were created specifically for special education research and practice. Given that the systematic review aims to examine research related to IEP quality, a legal document specific to special education, the CEC standards address the unique challenges of conducting research within this field. Moreover, these standards were developed collaboratively with special education researchers and practitioners, ensuring they reflect the priorities and realities of working and conducting research in this field (Council for Exceptional Children, 2014). Finally, the CEC standards align more closely with the practices of professional

organizations in special education (e.g., targeted outcomes based on the learner population). The decision to adopt the CEC standards was made to ensure that the evaluation of evidence-based practices is grounded in criteria specifically designed for the applied field of special education.

Each QED methodological rigor indicator set forth by CEC were used to score each primary study. The indicators are as follows: (1) context and setting, (2) participants, (3) intervention agent, (4) description of practice, (5) implementation fidelity, (6) internal validity, (7) outcome measures/dependent variable, and (8) data analysis. A total of 24 indicators were coded as some of the original eight indicators have sub-categories. The indicators, operational definitions, examples, non-examples, and supportive research literature were used to design a code book for the first author and research assistant to reference and make informed decisions about whether the primary studies meet each criterion. The first author designed a Qualtrics survey to code methodological rigor characteristics. The first author and research assistant coded each primary article independently. For each CEC indicator, the researchers coded either a “1,” meaning the article met all required components of the indicator, or a “0,” meaning the article did not meet all required components. The 24 indicators comprising the eight major and 16 subcategories were coded with either a zero or one. The full coding manual is provided in the supplemental materials.

Coding Interrater Reliability

The first author and research assistant independently coded all primary articles. The research assistant underwent one hour of training with the first author, where they were provided with a detailed code book that corresponds with the Qualtrics survey used to collect coded data. After the initial training, the first author and research assistant will independently code the same primary article and reconvene to discuss any discrepancies. After coding the study

characteristics, interrater reliability for the total agreement was 99.3% and 100% for the unscored agreement. The first author and research assistant reviewed the individual coding results, discussed any disagreements, and agreed to a consensus.

All data collected on methodological rigor were input into the Quality Matrix Tool created by Lane and associates (2014), to calculate the IRR. IRR for methodological rigor of the first article, before consensus was 82.14% and 100% after consensus. All remaining primary articles IRR for methodological rigor was 100% before consensus, reaching an overall IRR of 98.2%. Table 2 provides a breakdown of the IRR percentages for each methodological rigor indicator, prior to consensus.

Table 2

IRR for Individual Methodological Rigor Indicator

Methodological Rigor Indicator	IRR Percentage
1.0 Context and Setting	100%
2.0 Participants	94.45%
3.0 Intervention Agent	100%
4.0 Description of Practice	94.45%
5.0 Implementation Fidelity	100%
6.0 Internal Validity	96.3%
7.0 Outcome Measures	98.15%
8.0 Data Analysis	100%

Data Analysis

All data collected during the coding process underwent descriptive analysis using several methods. Categorical data, such as participant gender or study design, were analyzed through percentages and frequencies (George & Mallery, 2016). These calculations can be performed manually or with R using the 'Rscript-cat' and 'Chi-squared test in R' (Navarro & Kothe, 2019). When calculating manually, percentages can be obtained by dividing the number of items in a category by the total number of items and multiplying the result by 100. Percentages provide a

standardized numerical way to compare proportions across various sample sizes and studies (George & Mallery, 2016). This approach enabled the author to summarize and compare the prevalence of characteristics or outcomes across studies. Frequencies represent simple counts of how often a particular category or event occurs within the data set. Frequencies provided the researcher with a raw number that is useful for understanding the actual count of occurrences (George & Mallery, 2016).

Continuous data points, such as intervention effects or participant age, were analyzed by calculating means, standard deviations (SD), and interquartile ranges (IQR). The mean is an average of a set of numbers, calculated by adding all values and dividing by the total number of data points. The mean measures central tendency, providing a broad view of the dataset's values (Wan et al., 2014). This measure allowed the researcher to compare average intervention effects or descriptive characteristics across studies. A standard deviation (SD) measures the variability within a dataset, indicating how much data points deviate from the mean. SDs provided insights into how far each score differs from the mean, which is crucial for understanding the consistency of results across studies. Lastly, the interquartile range (IQR) is the difference between the 75th and 25th percentiles of the dataset, representing the middle 50% of the data (Wan et al., 2014). The formula for calculating the IQR is $IQR = Q3 - Q1$. This measure allowed the researcher to evaluate variability, offering a broader view when the data are not normally distributed or when outliers are present (Larson, 2006).

Results

Nine primary articles that satisfied all inclusion criteria met study inclusion criteria, as outlined in Table 3. The review results are structured by research question and discuss the observed trends. The results focus on the generalizability and feasibility of the interventions. All nine primary articles utilized group designs, with 55.56% ($n=5$) of articles using a pre- and post-

intervention design and 44.44% ($n=4$) using randomized controlled trials (RCT). All primary articles reported the type of IEP quality rubric used: 55.56% ($n=5$) utilized a rubric created by the researchers, 22.22% ($n=2$) used the R-GORI, 11.11% ($n=1$) used the Transition Requirements Checklist, and 11.11% ($n=1$) used the IQUIS. All studies that utilized researcher-created rubrics did not provide access to the full rubric or specific information regarding how to obtain access. This outcome is not ideal, as access to researcher-created rubrics is pivotal in the generalizability and ability to replicate the results.

Table 3

Primary Article Overview

Citation	Focus	Study Design	Quality Rubric
Burns et al., 2023	The effects of ongoing support on PLAAFP quality and student outcomes.	Pre/Post	Researcher Created
Doren et al., 2012	The effects of professional development on the quality of IEP goals.	Pre/Post	Researcher Created
Finn & Kohler, 2009	The effects of the Transition Outcome Project on the quality of postsecondary goals.	Pre/Post	Transition Requirements Checklist
Flannery et al., 2015	The effects of professional development on the quality of IEP transition components.	Pre/Post	Researcher Created
Lowman, 2016	The effects of three professional development models on the quality of standards-based IEP goals and objectives.	RCT	Researcher Created
Rakap & Balikci, 2024	The effects of ChatGPT on the quality of preschool IEP goals and objectives.	RCT	R-GORI

Rakap, 2023	The effects of ChatGPT on the quality of IEP goals and objectives written by novice special educators.	RCT	R-GORI
Ruble & McGrew, 2013	The effects of COMPASS on IEP goal quality and student outcomes.	RCT	Researcher Created
Shriner et al., 2012	The effects of a web-based decision-making tool on IEP goal quality.	Pre/Post	IQUIS–Goals/Objectives

Note: IQUIS = IEP Quality Indicator Scale for Goals/Objectives, RCT = randomized

controlled trial, R-GORI = Revised IEP/IFSP Goals and Objectives Rating Instrument.

Research Question 1: Study Characteristics

The study characteristics coding revealed patterns in the funding statements, authors, and publishing journals. The most common funding sources were grants from the Institute of Educational Sciences (IES) and the Office of Special Education Programs (OSEP), with four of the primary articles utilizing such funding (44.4% of all studies, 66.67% of all studies that reported funding). Three primary articles, 33.3%, reported no funding. In contrast, two articles reported grants from other sources, such as the National Institute of Mental Health (Ruble & McGrew, 2013) and the Illinois State Board of Education (Shriner et al., 2012). For article authorship, there were 27 authors across all nine primary articles. Four were authors or co-authors of two primary articles: Allison Lombardi, PhD; K. Brigid Flannery, PhD; Mimi McGrath Kato, MS; Salih Rakap, PhD. Articles appeared in 7 academic journals. *The Journal of Autism and Developmental Disorders* and *Career Development for Exceptional Individuals* both published two of the nine primary articles. Information on the publishing journals and reported funding are detailed in Table 4.

Table 4

Primary Article Journal and Funding Overview

Citation	Academic Journal	Funding
Burns et al., 2023	<i>Journal of Research in Special Educational Needs</i>	Private Funding
Doren et al., 2012	<i>Remedial and Special Education</i>	IES Grant
Finn & Kohler, 2009	<i>Career Development for Exceptional Individuals</i>	No Funding Reported
Flannery et al., 2015	<i>Career Development for Exceptional Individuals</i>	IES Grant
Lowman, 2016	<i>Communication Disorders Quarterly</i>	OSEP Grant
Rakap & Balikci, 2024	<i>Journal of Autism and Developmental Disorders</i>	No Funding Reported
Rakap, 2023	<i>Journal of Special Education Technology</i>	No Funding Reported
Ruble & McGrew, 2013	<i>Journal of Autism and Developmental Disorders</i>	National Institute of Mental Health Grant
Shriner et al., 2012	<i>The Journal of Special Education</i>	IES Grant & Illinois State Board of Education Grant

Note: IES = Institute of Education Sciences, OSEP = Office of Special Education

Programs

Information on the grade levels taught by participants was also collected: 11.1% ($n=1$) did not report this information, 33.3% ($n=3$) indicated elementary grade levels (pre-kindergarten through sixth grade; Burns et al., 2023; Rakap & Balikci, 2024; Shriner et al., 2012), and 55.6% ($n=5$) reported secondary and post-secondary grade levels (ninth grade through adulthood; Burns et al., 2023; Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015; Shriner et al., 2012). The results for grade levels taught showed a majority percentage for secondary and post-

secondary grade levels, which was expected as it correlates to the changes that occurred during the 2004 reauthorization of IDEA, highlighting the need for more detailed transition components.

Additional data on participant characteristics, such as gender, age, years of experience, and highest education level, were gathered. Although two articles did not disclose participant gender, the remaining primary articles indicated that the majority of participants identified as female (Burns et al., 2023; Doren et al., 2012; Flannery et al., 2015; Lowman, 2016; Rakap, 2023; Rakap & Balikci, 2024; Shriner et al., 2012). Similarly, two articles did not report years of experience (Finn & Kohler, 2009; Ruble & McGrew, 2012); however, six of the articles reported years of experience ranging from 8 to 15 years, while one reported an average of 14 months. The average years of experience across all six primary articles that reported the demographic was 9.67. The standard deviation (SD) was 4.3 years of experience, meaning that, on average, the participants' years of experience deviated by 4.3 years from the mean of 9.67 years. Additionally, the interquartile range (IQR) for the demographic was 5.4 years, indicating that the middle 50% of the participants' ages are spread across a range of 5.4 years. The IQR accounts for outliers, such as the article that reported an average of 14 months of experience, which is significantly lower than the averages reported by the other articles.

Regarding participant age, five of the nine primary articles (55.56%) discussed the demographic details, reporting average ages ranging from 24.5 to 42.8 years (Doren et al., 2012; Flannery et al., 2015; Lowman, 2016; Rakap, 2023; Rakap & Balikci, 2024). The average age of participants across all five articles that included this information was 37.45 years. The standard deviation (SD) was 7.4 years, indicating that, on average, participants' ages varied by 7.4 years from the mean of 37.45. The SD suggests a higher level of variability, and while the interquartile range (IQR) can address outliers and offer a more accurate perspective on the range, it could not

be calculated due to insufficient data in the primary articles. Finally, concerning the highest education level, only four primary articles (44.4%) reported on education (Burns et al., 2023; Doren et al., 2012; Rakap, 2023; Rakap & Balikci, 2024). Among those, the percentage of participants with a graduate degree varied from 0.09% to 77.78%, with bachelor's degrees being largely unreported. This outcome could be attributed to the nature of the participants, as speech-language pathologists are required to have a master's degree, and special educators must hold a bachelor's degree.

Study site location descriptors were the least reported study demographic, with six articles reporting a geographical descriptor (Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015; Lowman, 2016; Ruble & McGrew, 2013; Shriner et al., 2012) and two reporting location type (e.g., rural, suburban, urban; Flannery et al., 2015; Shriner et al., 2012).

Geographical descriptors were most commonly in the form of cardinal direction and country (e.g., Northwestern United States, Midwestern state). For location type, one article reported suburban and rural, and the other reported urban, suburban, and rural. Additional information pertaining to the number of participants per location type was not reported. These findings are reasonable, as location descriptors can directly impact the level of confidentiality (Owings et al., 2025).

All nine primary articles reported the IDEA disability category for the submitted IEPs used in the study. The most frequently reported disability category was Autism, with eight of the articles identifying this category, representing 88.89% ($n=8$) of all studies. Five of the articles indicated more than one disability category (55.56% of all studies; Burns et al., 2023; Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015; Shriner et al., 2012). Four articles identified only one disability category: speech-language impairment (SLI; Lowman, 2016) and

autism (ASD; Rakap, 2023; Rakap & Balikci, 2024; Ruble & McGrew, 2013). A total of ten disability categories were reported across all nine primary articles. The reported percentages are as follows: autism (ASD; 88.89% of articles, $n=8$), specific learning disability (SLD; 55.56% of articles, $n=5$), emotional disturbance (ED; 44.44% of articles, $n=4$), intellectual disability (ID; 55.56% of articles, $n=5$), other health impairment (OHI; 44.44% of articles, $n=4$), speech-language impairment (SLI; 44.44% of articles, $n=4$), vision impairment (VI; 22.22% of articles, $n=2$), hearing impairment (HI; 11.11% of articles, $n=1$), multiple disabilities (MD; 11.11% of articles, $n=1$), orthopedic impairment (OI; 11.11% of articles, $n=1$), and traumatic brain injury (TBI; 11.11% of articles, $n=1$). Lastly, all primary articles reported the number of IEPs reviewed during the study. The average number of IEPs across all studies was 269.22, with a standard deviation of 136.11, indicating that, on average, the number of IEPs reviewed per study deviated from the mean by 136.11 IEPs. When calculated using IQR, the middle 50% of the number of IEPs reviewed across all studies averaged 248. Table 5 provides an overview of the disability categories reported and the number of IEPs reviewed for each article.

Table 5

Disability Category and Number of IEPs Reviewed

Citation	Disability Category	Number of IEPs Reviewed
Burns et al., 2023	Autism, ED, ID, OHI, SLD	211
Doren et al., 2012	Autism, ID, SLD	137
Finn & Kohler, 2009	Autism, ED, HI, ID, MD, OHI, SLD, SLI, VI	463
Flannery et al., 2015	Autism, ED, ID, OHI, OI, SLD, SLI, TBI, VI	302
Lowman, 2016	SLI	147

Rakap & Balikci, 2024	Autism	330
Rakap, 2023	Autism	450
Ruble & McGrew, 2013	Autism	79
Shriner et al., 2012	Autism, ED, ID, OHI, SLD, SLI	304

Note: ED=emotional disturbance, ID=intellectual disability, HI=hearing impairment, MD=multiple disabilities, OHI=other health impairment, SLD=specific learning disability, SLI=speech language impairment, TBI=traumatic brain injury, and VI=vision impairment

Research Question 2: Intervention(s)

Although each article employed different interventions to enhance IEP quality, professional development models were the most common independent variable (44.4%, $n = 4$; Burns et al., 2023; Doren et al., 2012; Flannery et al., 2015; Lowman, 2016). Three primary articles utilized web-based programs as the independent variable (Rakap, 2023; Rakap & Balikci, 2024; Shriner et al., 2013). While most articles featured professional development models as the independent variable, three articles included training or professional development aimed at equipping participants with skills to use independent variables such as IEP development models (Ruble & McGrew, 2013) and web-based tools (Finn & Kohler, 2009; Shriner et al., 2012). Each professional development model varied in duration, content, and delivery method. At a minimum, some models involved a one-day training session (Burns et al., 2023), while at a maximum, they extended across an entire academic year (Doren et al., 2012; Flannery et al., 2015). All nine primary articles analyzed the quality of one or more components of IEPs. Instead of using case study examples, all articles utilized written and submitted IEP components from participants. Information about the cost of each intervention was absent from all of the primary

articles. Given that the systematic review focused on the feasibility and generalization of the independent variable, the lack of cost information is concerning. Table 6 provides additional information on the independent variables and training length when applicable for each primary article.

Table 6

Primary Article Independent Variable(s)

Citation	Independent Variable(s)	Training Length
Burns et al., 2023	Ongoing support using the online Goal Book Toolkit versus a singular lecture style training	1-day for lecture style training
Doren et al., 2012	Professional Development Model	One academic year (4 half-day sessions and 90 minutes of coaching sessions)
Finn & Kohler, 2009	Transition Outcomes Project	Not Specified
Flannery et al., 2015	Professional Development Model	Two full days of PD and two 1.5 days and four 90-minute professional learning community (PLC) meetings
Lowman, 2016	Three Professional Development Models	Workshop Only (4 Hours); Peer Coaching Group (5 Hours and 2-Months of Coaching Meetings; Web-Based Only Group (Not Applicable)
Rakap & Balikci, 2024	ChatGPT	Not Applicable
Rakap, 2023	ChatGPT	Not Applicable
Ruble & McGrew, 2013	Collaborative Model for Promoting Competence and Success (COMPASS)	Initial Consultation; Coaching sessions every 4-6 weeks
Shriner et al., 2012	Web-Based Individualized Education Program Decision-Making Tutorial	Initial Training (3 Hours); Access to ongoing support

Research Question 3: Intervention Component(s) Effectiveness

For the studies that utilized professional development models, the results on the effectiveness of intervention components vary. Burns and colleagues (2023) found that the one-day training alone did not improve the quality of PLAAFPS written by the participants; however, continued access to the online resources was a significant predictor of PLAAFP quality. The results also revealed higher access rates to online resources were associated with higher quality scores, particularly when the resources were accessed for 14 hours or more over one academic year (Burns et al., 2023). Doren and associates (2012) discovered that participating in the PD model significantly predicted improved postsecondary goal quality. The researchers also utilized hierarchical linear modeling (HLM) and moderation analysis to determine correlations between participant characteristics and levels of quality improvement. Participant characteristics, including educator placement, caseload size, and years of experience, were identified as predictors of goal quality. For instance, participants who taught in self-contained classrooms demonstrated higher levels of goal quality throughout the study (Doren et al., 2012).

Additionally, it was determined that the amount of professional development the participants engaged in was not a predictive factor for goal quality. Flannery and colleagues (2015) found that a professional development model increased the quality of all analyzed IEP components, including PLAAFPs, post-secondary goals, and course of study. Specifically, the “impact of disability” portion of the PLAAFP statements revealed higher improvement rates than the other PLAAFP components. Lowman (2016) employed three professional development models: web-based, workshop-only, and peer-coaching. The workshop-only and the peer-coaching groups yielded higher improvement rates than the web-based group. The authors noted

that the online portion used by the web-based group was an effective mode of imparting knowledge but did not affect goal quality (Lowman, 2016).

In studies that did not employ professional development models, the levels of effectiveness varied. Rakap (2023) and Rakap and Balikci (2024) used ChatGPT to enhance the quality of IEP goals. Rakap (2023) found that participants who had received previous training on goal quality had higher improvement rates than those who had not. Rakap and Balikci (2024) discovered that using ChatGPT increased goal quality. It was also noted that, through ChatGPT, participants began writing goals related to social skills, motor, sensory, and self-care. In contrast, the control group's goals typically pertained solely to academic objectives.

Three primary studies employed interventions that involved training participants (Finn & Kohler, 2009; Ruble & McGrew, 2013; Shriner et al., 2012). Finn and Kohler (2009) found that using the *Transition Outcomes Project*, a web-based tool, improved the overall quality of IEP transition components. However, clear communication during participant training and ongoing web-based tool support also led to higher improvement rates. Ruble and McGrew (2013) discovered that collaborative treatment planning, used in the COMPASS model, improved IEP quality rates. Participants engaged in coaching sessions that included feedback and self-reflection. The results revealed that student involvement significantly predicted higher levels of IEP quality (Ruble & McGrew, 2013). Shriner and associates (2012) provided participants three hours of training and continued human support for using a web-based decision-making tool, the *IEP Quality Tutorial*. Three participant groups existed: a control, partial, and full-use group. The control group, which submitted IEP goals, showed no change in quality scores. Both the partial-use group, which minimally accessed the online tool, and the full-use group, which continually accessed the online tool, yielded increased quality scores (Shriner et al., 2012). Both groups with

access to the online tool received three hours of training on using the *IEP Quality Tutorial*.

However, the group with full access and continued tool use had higher improvement rates than the other two groups.

Overall, the interventions for all studies positively affected IEP quality. Ongoing support was the main intervention component that yielded higher improvement rates across multiple studies and was documented through either the amount of time spent accessing online materials or the number of in-person support sessions. Ongoing support was delivered in several different modalities through continued access to materials (Burns et al., 2023), peer coaching (Lowman, 2016), professional learning communities (Flannery et al., 2015), access to continued human support (Finn & Kohler, 2009), and coaching sessions (Doren et al., 2012; Ruble & McGrew, 2013). These results reveal that when intervening in IEP quality, singular training, despite length, format, and content, is less effective if not paired with continued support.

Research Question 4: Addressing Concerns

Regarding the extent to which the studies addressed concerns about IEP quality found in the literature, all interventions did so. The literature indicates that the most common IEP components with quality issues are the PLAAFP statements and annual goals (Hott et al., 2020; Hott et al., 2021; Klang et al., 2016). PLAAFP statements are often misaligned or incomplete (Hott et al., 2020), while annual goals are frequently unmeasurable, unclear, and misaligned (Hott et al., 2021; Klang et al., 2016). The dependent variable for most primary articles was the quality of IEP goals and objectives (88.89% of all studies, $n=8$; Dorne et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015; Lowman, 2016; Rakap, 2023; Rakap & Balikci, 2024; Ruble & McGrew, 2023; Shriner et al., 2012). Three studies analyzed the quality of multiple IEP components, specifically the IEP components related to the topic of transition (33.33% of all

studies, $n=3$); Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015). These results are relevant to the timeline of IDEA changes, as the most recent reauthorization of the law instated significant changes for transition components and the quality of IEP goals (Individuals with Disabilities Education Improvement Act, 2004). Additionally, three studies addressed the quality of PLAAFP statements; one focused on general quality (Burns et al., 2023), and two on the quality of transition components (Finn & Kohler, 2009; Flannery et al., 2015). With PLAAFP quality being a major concern, it is alarming that only one article (Burns et al., 2023) attempted to intervene on overall PLAAFP quality that was not tied to transition components. On the other hand, most primary articles focused on the quality of goals and objectives.

Research Question 5: Methodological Rigor

Of the 24 indicators and subcategories of methodological rigor, 11 were met 100% of the time by all primary articles, representing 45.83% of all coded indicators. The 11 met indicators include 100% of indicators for section 1.0 for reporting context and setting, 50% of section 2.0 for reporting participant demographics, 100% of section 4.0 for reporting critical features of the intervention, 33.33% of section 6.0 for reporting information regarding internal validity, 66.67% of section 7.0 for the appropriateness of the outcome measures used, and 50% of section 8.0 for the appropriateness of the data analysis used. Indicators 6.8, reporting on or having low attrition across groups, and 6.9, reporting on or having low differential attrition between groups, had the lowest percentage of being met by all articles, with only 11.11% ($n=1$), of articles meeting the indicators (Finn & Kohler, 2009). Subsequently, indicator 7.4, analyzing the appropriateness of the frequency and timing of the outcome measures, was met by 44.44% ($n=4$) of articles (Finn & Kohler, 2009; Flannery et al., 2015; Lowman, 2016; Shriner et al., 2012). Table 7 provides the percentages of articles that met each quality indicator.

Table 7*Percentages of Indicators Met Across All Articles*

Methodological Rigor Indicator	Percentage of Articles
1.0 Context and Setting	100%
1.1 Context and Setting	100%
2.0 Participants	94.45%
2.1 Participants	100%
2.2 Participants	88.89%
3.0 Intervention Agent	66.67%
3.1 Intervention Agent	77.78%
3.2 Intervention Agent	55.56%
4.0 Description of Practice	100%
4.1 Description of Practice	100%
4.2 Description of Practice	100%
5.0 Implementation Fidelity	77.78%
5.1 Implementation Fidelity	66.67%
5.2 Implementation Fidelity	88.89%
5.3 Implementation Fidelity	77.78%
6.0 Internal Validity	57.41%
6.1 Internal Validity	100%
6.2 Internal Validity	100%
6.3 Internal Validity	66.67%
6.4 Internal Validity	55.56%
6.8 Internal Validity	11.11%
6.9 Internal Validity	11.11%
7.0 Outcome Measures	87.04%
7.1 Outcome Measures	100%
7.2 Outcome Measures	100%
7.3 Outcome Measures	100%
7.4 Outcome Measures	44.44%
7.5 Outcome Measures	100%
7.6 Outcome Measures	77.78%
8.0 Data Analysis	83.34%
8.1 Data Analysis	100%
8.3 Data Analysis	66.67%

As for the overall number of indicators, only one article met enough indicators without being weighted to meet the CEC standards for being considered an evidence-based practice (Shriner et al., 2012). The weighted number of indicators met accounts for all sub-indicators met and does not require that all sub-indicators be met for a section to be counted. After the 80% weight, six articles, 66.67% of the total primary articles, met enough quality indicators to be considered an evidence-based practice per the CEC standards (Burns et al., 2023; Finn & Kohler, 2009; Flannery et al., 2015; Rakap, 2023; Rakap & Balikci, 2024; Shriner et al., 2012).

Discussion

The quality of Individual Education Programs (IEPs) plays a vital role in students' ability to make meaningful progress and access the general education curriculum and rights afforded to students with disabilities (Harmon et al., 2020). Although the quality of IEP components has been analyzed (Burns et al., 2023; Hott et al., 2021; McKenna et al., 2023), results of this systematic review reveal that very few empirical studies have been conducted to improve IEP quality. Many factors contribute to the lack of IEP quality, such as the nationwide educator shortage, disproportionately affecting high-needs areas and special education (Peyton & Acosta, 2022; Schmitt & DeCourcy, 2022). Even though the current state of public education has caused uncertainty for many, students with disabilities continue to be protected under federal law and are, by law, entitled to a Free and Appropriate Public Education (FAPE; IDEA, 2004), where IEPs are a critical pillar.

In the systematic review, nine articles met the inclusion criteria and focused on actionable research to improve the quality of one or more IEP components. The included articles primarily concentrated on enhancing the quality of IEP goals and objectives, a frequently cited concern throughout the literature. The lack of diversity across the addressed IEP components is troubling

since each IEP contains numerous elements beyond annual goals and objectives (Harrison et al., 2013; Joyce et al., 2018). For instance, only one primary article included PLAAFP statements, the introduction, and supporting statements for all decisions made within the IEP (Goran et al., 2020). Although several articles examined PLAAFP quality, they focused solely on its relationship to transition components. All IEPs are required to include PLAAFP statements, as this component is not specific to transition-aged students.

The results of this systematic review raise concerns about methodological rigor and the reporting of intervention details. Considering only one article fulfilled the criteria for being an established evidence-based practice, the scarcity of evidence-based practices accessible to educators and school administrators is concerning. For the articles that did meet the standards for evidence-based practices, once weighted, the insufficient reporting on the length and cost of the intervention, the role and training of the intervention agents, and the lack of access to any resources that determine intervention quality, makes it challenging for the studies to be replicated for implementation by school staff.

Implications for Practice

This systematic review benefits researchers and practitioners. First, there is a common thread across many primary articles. Access to ongoing support in various forms can positively impact the continued growth of participating educators (Burns et al., 2023; Doren et al., 2012). In the past, professional development for educators has primarily consisted of singular training sessions, which can sometimes be lengthy and costly (Donath et al., 2023). Although the findings showed that online formats are effective for imparting knowledge, singular training without follow-up is less effective than those that include ongoing support after the content is delivered to participants (Lowman, 2016; Shriner et al., 2012). The research indicates that this ongoing

support does not need to be costly or extensive, as many studies offered participants continued access to online resources and support (Burns et al., 2023; Lowman, 2016; Shriner et al., 2012). Research also demonstrates that access to professional learning communities (PLCs) can yield positive results, as peer coaching is a valuable tool for educators and students (Flannery et al., 2015; Lowman, 2016).

Limitations

Some limitations should be discussed regarding this systematic review. One limitation is that only studies involving an intervention were included. Numerous descriptive reviews provide critical information and context related to the issue of IEP quality. Further research should be conducted to give readers a comprehensive overview of the current literature on IEP quality. Additionally, the timeframe from which articles were selected was limited to those published from 2004 to the present. This was initially based on the most recent reauthorization of IDEA in 2004, but a broader search including articles from before 2004 may benefit researchers and readers. Although IDEA saw several changes during the 2004 reauthorization, many components remained relatively unchanged. Those areas may contain IEP quality research conducted before 2004, potentially offering additional interventions not included in this review. Lastly, the current review utilized CEC methodological rigor standards instead of the commonly used What Works Clearinghouse (2022) standards for group design studies. All reviewed articles were considered group design studies, and the results may have benefited from employing WWC standards. Additional analysis using the WWC standards may prove advantageous for researchers and readers.

Future Research

Future research on IEP quality will greatly benefit researchers and practitioners if it focuses on generalizing interventions across settings, grade levels, and demographics. To effectively address IEP quality, a well-documented concern for public schools, interventions must meet methodological rigor standards and report the information specified by those standards. Practitioners cannot adequately replicate and implement interventions in their daily work without comprehensive information. Additionally, future research could explore whether providing participants with the rubrics used to assess quality can enable them to self-monitor their IEP quality and achieve a positive impact overall. Self-monitoring is typically used alongside other interventions, such as professional development models, goal setting, or positive reinforcement (Cooper et al., 2020). However, with or without additional intervention components, self-monitoring may be a feasible and cost-effective option for improving IEP quality. Lastly, identifying and utilizing a widely recognized training model could enhance the applicability of study results for practitioners. Models like Behavior Skills Training (BST; Kirkpatrick et al., 2019), which is commonly used in Applied Behavior Analysis, are becoming increasingly prevalent in educational research. The Professional Learning Model (PLM; Courtney, 2018) is a formal training approach that helps trainers and PD hosts develop and implement professional development opportunities tailored to their participants. Utilizing BST, PLM, or other established training models can benefit practitioners, as the well-documented details of these interventions can be easily applied in their everyday work.

Conclusion

Overall, the results of the systematic review revealed a significant gap in the literature regarding empirically validated research and evidence-based practices for improving IEP quality. Since IEP quality is directly related to positive academic and behavioral outcomes for students

with disabilities, further research is essential in this area (Srivastava, 2024; Talbott et al., 2017).

Future studies must also consider replicability and generalizability during the planning and execution phases. Although the articles reviewed in the systematic review indicated that the interventions positively impacted IEP quality, the lack of detail in the published work complicates the task for future researchers or practitioners attempting to implement the reported interventions.

References

- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cooper, J., Heron, T., & Heward, W. (2020). *Applied behavior analysis* (3rd ed.). Pearson.
- Council for Exceptional Children. (2014). *Exceptional Children*, 80(4), 504–511.
<https://doi.org/10.1177/0014402914531388>
- Courtney, S. A. (2018). Teacher educator-embedded professional learning model. *International Electronic Journal of Mathematics Education*, 13(3). <https://doi.org/10.12973/iejme/2702>
- Doren, B., Flannery, K. B., Lombardi, A. R., & Kato, M. M. (2012). The impact of professional development and student and teacher characteristics on the quality of postsecondary goals. *Remedial and Special Education*, 34(4), 215–224.
<https://doi.org/10.1177/0741932512468037>
- Finn, J. E., & Kohler, P. D. (2008). A compliance evaluation of the transition outcomes project. *Career Development for Exceptional Individuals*, 32(1), 17–29.
<https://doi.org/10.1177/0885728808315332>
- George, D., & Mallery, P. (2016). *IBM SPSS Statistics 23 Step by Step*. Routledge.
- Goodwin, K., Farquharson, K., Yeager Pelatti, C., Schneider-Cline, W., Harvey, J., & Bush, E. (2020). Examining the quality of individualized education plan (IEP) goals for children with traumatic brain injury (TBI). *Communication Disorders Quarterly*, 43(2), 152574012097611. <https://doi.org/10.1177/1525740120976113>
- Hott, B. L., Brigham, F. J., & Peltier, C. (2021a). *Research methods in special education* (pp. 99–100). Slack Incorporated.

- Hott, B. L., Brigham, F. J., & Peltier, C. (2021b). *Research methods in special education* (pp. 89–111). Slack Incorporated.
- Hott, B. L., Jones, B. A., Randolph, K. M., Kuntz, E., McKenna, J. W., & Brigham, F. J. (2020). Lessons learned from a descriptive review of rural individualized education programs. *The Journal of Special Education*, 55(3), 002246692097267. <https://doi.org/10.1177/0022466920972670>
- Hott, B. L., Morano, S., Peltier, C., Pulos, J., & Peltier, T. (2020). Are students with mathematics learning disabilities receiving FAPE?: Insights from a descriptive review of individualized education programs. *Learning Disabilities Research & Practice*, 35(4), 170–179. <https://doi.org/10.1111/ldrp.12231>
- Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446, 118 Stat. 2647 (2004)
- kappa2: Cohen's Kappa and weighted Kappa for two raters in irr: Various Coefficients of Interrater Reliability and Agreement. (n.d.). Rdrr.io. Retrieved July 29, 2024, from <https://rdrr.io/cran/irr/man/kappa2.html>
- Kirkpatrick, M., Akers, J., & Rivera, G. (2019). Use of behavioral skills training with teachers: A systematic review. *Journal of Behavioral Education*, 28(3), 344–361. <https://doi.org/10.1007/s10864-019-09322-z>
- Klang, N., Rowland, C., Fried-Oken, M., Steiner, S., Granlund, M., & Adolfsson, M. (2016). The content of goals in individual educational programs for students with complex communication needs. *Augmentative and Alternative Communication*, 32(1), 41–48. <https://doi.org/10.3109/07434618.2015.1134654>

- Kurth, J., & Mastergeorge, A. M. (2009). Individual education plan goals and services for adolescents with autism: Impact of age and educational setting. *The Journal of Special Education*, 44(3), 146–160. <https://doi.org/10.1177/0022466908329825>
- Larson, M. G. (2006). Descriptive statistics and graphical displays. *Circulation*, 114(1), 76–81. <https://doi.org/10.1161/circulationaha.105.584474>
- Lowman, J. J. (2015). A comparison of three professional development mechanisms for improving the quality of standards-based IEP objectives. *Communication Disorders Quarterly*, 37(4), 211–224. <https://doi.org/10.1177/1525740115595338>
- McKenna, J. W., Solis, M., Garwood, J., & Parenti, M. (2023). Characteristics of individualized education programs for students with learning disabilities: A systematic review. *Learning Disability Quarterly*, 1. <https://doi.org/10.1177/07319487231182697>
- Munn, Z., Peters, M.D.J., Stern, C. *et al.* Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 18, 143 (2018). <https://doi-org.ezproxy.lib.ou.edu/10.1186/s12874-018-0611-x>
- National Center for Education Statistics. (2024). Students with Disabilities. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved May 30, 2024, from <https://nces.ed.gov/programs/coe/indicator/cgg>.
- Navarro, D., & Kothe, E. (2019). Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1). In *learningstatisticswithr.com*. <https://learningstatisticswithr.com/book/index.html>
- Owings, J., Barwick, A., & Smith, W. (2025). Chapter 2, meeting the challenges of data collection/safety in numbers: Collecting and using crime, violence and discipline incident

data to make a difference in schools. Ed.gov.

<https://nces.ed.gov/pubs2002/safety/chapter2.asp?t>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D.,

Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J.,

Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E.,

McDonald, S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: an Updated

Guideline for Reporting Systematic Reviews. *British Medical Journal*, 372(71).

<https://doi.org/10.1136/bmj.n71>

Ruble, L. A., McGrew, J., Dalrymple, N., & Jung, L. A. (2010). Examining the quality of IEPs

for young children with autism. *Journal of Autism and Developmental Disorders*, 40(12),

1459–1470. <https://doi.org/10.1007/s10803-010-1003-1>

Srivastava, R. (2024). The influence of individualized education programs (IEPs) on student

achievement. *Global International Research Thoughts*, 12(1).

<https://doi.org/10.36676/girt.v12.i1.03>

Talbott, E., Maggin, D. M., Van Acker, E. Y., & Kumm, S. (2017). Quality Indicators for

Reviews of Research in Special Education. *Exceptionality*, 26(4), 245–265.

<https://doi.org/10.1080/09362835.2017.1283625>

TMC Library. (2020). *LibGuides: Systematic Reviews: Citation Searching*. Tmc.edu.

https://libguides.library.tmc.edu/Systematic_Reviews/citationsearching?t

Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. (2008). Practical Implications of Data

Reliability and Treatment Integrity Monitoring. *Behavior Analysis in Practice*, 1(2), 4–

11. <https://doi.org/10.1007/bf03391722>

- Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, 14(1). <https://doi.org/10.1186/1471-2288-14-135>
- What works clearinghouse procedures and standards handbook, version 5.0 A publication of the national center for education evaluation at IES.* (2022). Institute of Educational Sciences. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf
- Yell, M., Katsiyannis, A., Ennis, R. P., & Losinski, M. (2013). Avoiding procedural errors in individualized education program development. *TEACHING Exceptional Children*, 46(1), 56–64. <https://doi.org/10.1177/004005991304600107>

Article 2

IEP PLAAFP Development Training

Nearly 15 percent of all public school students-almost 7.2 million children receive special education and related services outlined in Individualized Education Programs (IEPs), making the quality of these plans more critical than ever for student success (U.S. Department of Education, Office of Special Education Programs, 2023). Individualized Education Programs (IEPs) are legal documents that summarize a student's current academic and behavioral progress and detail the supports needed to ensure the student receives a free and appropriate public education (FAPE) (IDEA, 2004). IEPs comprise several components that guide school staff, parents, and students in creating and implementing the necessary educational plans for students with disabilities. One essential component of an IEP is the Present Levels of Academic Achievement and Functional Performance (PLAAFP) statement. A PLAAFP statement provides the educational team with an overview of the student's disability and its impact on their education, as well as assessment data, strengths, areas of need, annual goals, accommodations, modifications, and special education and related services. The ruling in *Endrew v. Douglas County School District* (2017) determined that IEPs must be "reasonably calculated to enable a child to make progress appropriate in light of the child's circumstances," emphasizing the importance of PLAAFP statements. When PLAAFP statements are comprehensive, data-driven, and based on the student's current circumstances, educational teams can ensure that the supports chosen are necessary and appropriate for the student to make adequate progress toward their educational goals (Harmon et al., 2020).

Common PLAAFP Errors

Several common errors occur when developing PLAAFP statements, such as a lack of (1) accurate and current data (Hott et al., 2021), (2) recent and appropriate data reported in a

student's present level of performances (Bateman, 2011), and (3) disability impact statement (Yell et al., 2016). Hott and associates (2019) conducted a descriptive review of IEPs for students in rural districts with emotional and behavioral needs. The researchers analyzed 126 IEPs across five rural school districts, focusing on the prevalence of behavioral needs in the IEPs. Less than three-quarters (74%, $n = 70$) of the IEPs had PLAAFP statements and annual goals that addressed the student's behavioral needs. Additionally, 68% ($n = 65$) of the total PLAAFP statements relied solely on teacher observations or anecdotal teacher reports as the primary source of social and behavioral data. Hott and colleagues (2020) conducted a similar descriptive review of IEPs for students, not specifically those with behavioral needs, across seven rural school districts. The researchers examined 133 IEPs and evaluated the PLAAFP statements, IEP goals, progress monitoring, and the alignment between the PLAAFP statements and IEP goals. The results revealed that 74% ($n = 70$) of the IEPs did not include complete disability impact statements, and less than 7% ($n = 6$) of the IEPs included PLAAFP statements, annual goals, and progress monitoring methods that met IDEA standards (Hott et al., 2020).

PLAAFP statements often fail to meet the legal requirements of the IDEA and state regulations (Hott et al., 2019; Hott et al., 2020). Poorly constructed PLAAFP statements can and do impact student outcomes (Kurth & Mastergeorge, 2010; Ruble et al., 2010; Yell et al., 2016). Poor-quality IEPs and vague PLAAFP statements are associated with lower educational outcomes for the students (Ruble et al., 2010). Similarly, Yell and associates (2016) found that poorly constructed PLAAFP statements often fail to describe how a student's disability can affect their involvement in the general education setting, which can lead to inappropriate placement and service decisions. Kurth and Mastergeorge (2010) found similar results during a descriptive review of IEPs of students with autism. Their results revealed that IEPs with vague PLAAFPs

were commonly associated with decreased goal attainment. Improving quality is increasingly important given the documented impact of low-quality PLAAFP statements on student outcomes.

Intervening on PLAAFP Quality

Although there is research that intervenes on IEP components such as goals (Lowman, 2015; Rakap, 2023; Ruble & McGrew, 2013; Shriner et al., 2013) and transition components (Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015), there is little research that intervenes on the quality of PLAAFP statements (Burns et al., 2023). A significant portion of the literature encompasses professional development (PD) models, which encompass a wide range of training techniques used to develop new skills or enhance existing ones, thereby improving IEP quality (Burns et al., 2022; Flannery et al., 2015; Lowman, 2015). For example, Flannery and colleagues (2015) conducted a study involving special educators participating in a PD model aimed at improving the quality of IEPs for students who needed transition components. Although the primary focus of the research was on the transition components of the IEP, the researchers also analyzed the quality of the PLAAFP statements before and after attendance at the PD sessions. The PD model consisted of two full days of in-person training and six subsequent meetings, held over six weeks, each ranging from 90 minutes to one and a half days in duration. The researchers concluded that there was a significant improvement in one area of the PLAAFP statements after the PD concluded: the impact of disability statements. Before engaging in the PD, 64% of PLAAFP statements included a disability impact statement; after the PD, this increased to 88%. Burns and associates (2022) conducted a similar study to analyze the impact of a professional development (PD) model on the quality of PLAAFP statements written by special

educators. The participants were divided into two groups: one received access to an online toolkit that provided resources and guidance on writing effective PLAAFP statements.

In contrast, the second group had access to the online toolkit and participated in a one-day in-person training session. The results showed that participants who attended the one-day training did not experience significant improvements in the quality of their PLAAFP.

Additionally, having ongoing access to PD resources was associated with high PLAAFP quality (Burns et al., 2022). Across the literature, PD models improve IEP quality, whether focusing on specific goals, transition components, PLAAFP statements, or the IEP as a whole. Although PD can be an effective way to train educators and increase IEP quality, most of the literature base requires extensive time commitments on behalf of the participants (Burns et al., 2022; Flannery et al., 2015; Lowman, 2015).

Behavior Skills Training

Many PD models have been employed to increase IEP quality, but often require participants to attend full or multi-day trainings (Burns et al., 2022; Flannery et al., 2015; Lowman, 2015). With that said, studies revealed that PD models that include continued support yield improved results (Burns et al., 2022; Lowman, 2015). One coaching model commonly used in Applied Behavior Analysis (ABA) is the Behavior Skills Training (BST) (DiGennaro Reed et al., 2018; Hogan et al., 2014). BST follows a four-part process, consisting of instruction, modeling, rehearsal, and feedback, which is commonly used to assist individuals in acquiring skills (Flynn & Lo, 2016) and re-establishing previously learned skills (Miller et al., 2014). Flynn and Lo (2016) employed the BST coaching model to train educators in conducting functional analyses and implementing supplementary interventions. The training sessions were relatively short, lasting no more than 2 hours, and yielded positive results in the educator's skill

acquisition. The majority of the participants were able to generalize the skills learned to other cases. Van Vonderen and associates (2012) employed the BST coaching model to assess the effectiveness of the instruction step, as well as video feedback on correct response prompting by staff, when working with children. The results showed that combining the instruction phase and video modeling effectively increased the staff's correct prompting behavior. Lastly, Miller and colleagues (2014) used the BST coaching model as a "booster" training to re-establish classroom management skills for educators. The booster training included a review of five classroom management tools previously taught to educators in a BST-style professional development (PD) a year prior. The results showed that the booster session was effective in reestablishing the educator's classroom management skills, and improvements were seen in both the classroom and assessments conducted during the training session.

BST coaching model can be effective to support staff training and educational settings as it typically promotes short and time-efficient training sessions (Flynn & Lo, 2016); the ability to be conducted in-person or in-vivo is supported by research demonstrating the effectiveness of both approaches (van Vonderen et al., 2012); the ability to be used for acquiring new skills (Flynn & Lo, 2016); re-establishing previously learned skills (Miller et al., 2014); and research results often showing a high social validity outcome (Flynn & Lo, 2016; Hall et al., 2009; Miller et al., 2014). For these reasons, the independent variable will be designed to include several components the literature has shown to be effective in behavior change. The intervention package will be designed based on the BST coaching model, including video instruction and modeling (van Vonderen et al., 2012), continued support past the initial PD (Burns et al., 2022; Lowman, 2015), and all training and coaching sessions being designed with time efficiency in mind.

Purpose

This study aims to evaluate the efficacy of the BST coaching model in improving the quality of PLAAFP statements written by special educators in high-needs elementary schools. To decrease the time commitment for participants and maintain the integrity of the content included in the professional development (PD), the proposed study will utilize an online instruction and modeling video, as well as in-person coaching that follows the behavior skills training (BST) coaching model.

The research questions are as follows:

1. Is there a functional relation between the use of PD, utilizing the BST coaching model and in-person coaching, and increased quality of PLAAFP statements written by special educators in elementary schools?
2. How do special educators perceive the effectiveness and feasibility of the PLAAFP professional development with follow-up coaching?

Method

Setting

A public school is considered high-needs when it is either ranked in the top quartile of schools by the percentage of students from low-income families; it is located in an area where at least 30% of students come from families with incomes below the poverty line; or at least 60% of students are eligible for free or reduced-price school lunch (Elementary and Secondary Education Act, 1965). The elementary schools chosen to participate in the study were selected due to their existing working relationship with the researchers and both being classified as high-needs areas. Both head principals for the two locations met with the first author. They expressed

concerns about the quality and compliance of the IEPs written by special educators in their buildings, stating that they could benefit from more individualized training.

The first school, Johnson Elementary, is a Title 1 high-need elementary school in the south-central United States. The school has an enrollment of 450 students, with approximately 100 receiving special education services under one or more IDEA categories, which accounts for approximately 22% of the student population. The school staff included a head principal, a counselor, and four special education teachers. The second school, Abbott Elementary, is a Title I, high-needs elementary school located in the south-central United States. It serves approximately 400 students, with 70 receiving special education services under one or more IDEA categories, which accounts for roughly 17.5% of the student population. The school staff included a head principal, a counselor, and three special education teachers. High-needs schools are educational institutions that face significant challenges and require additional support to meet the needs of their students (Sleeter, 2001). Both settings were part of the same school district, therefore following the same district-specific special education guidelines. The school district serves roughly 28,000 students total, with around 6,000 students receiving special education services. The school district's special services department employs approximately 300 special educators.

Participants

The participants varied in demographics, such as age, years of teaching experience, and highest level of education. See Table 1 for a summary of participant demographics. To qualify for the study, each potential participant used the link in the recruitment email to review the consent document, give electronic consent, complete their demographic information, and then be prompted to email a de-identified PLAAFP statement written by them to a designated school

district employee, who will review it to ensure FERPA compliance and then submit it to the researchers. The researcher then reviewed and scored the submitted PLAAFP statement based on the SERC IEP quality rubric (2013). A quality percentage was calculated by dividing the points earned by the total number of points possible and then multiplying the result by 100. Participants who scored 75% or below were considered for participation in the study. The cut score was based on the current district educators' needs. Once consent was gained, baseline data collection for all participants began. All participants engaged in the same intervention components. Participants were grouped by their school location.

Table 1

Participant Demographics

Characteristic	Participant 1	Participant 2	Participant 3	Participant 4
Study Setting	School 1	School 1	School 1	School 2
Sex	Female	Female	Female	Female
Age Range	25-34	35-44	35-44	35-44
Ethnicity	White	White	White	White
Highest Level of Education	Bachelor's Degree	Master's Degree	Master's Degree	Bachelor's Degree
Years Teaching Special Education	1 Year	14 Years	10 Years	4 Years
Teaching Certificate Type	Alternative	Traditional	Alternative	Traditional
Grade Level(s) Taught	3 rd -6 th	3 rd -6 th	K-6 th	K-6 th
Current Caseload Size	21	19	12	17

Initial PLAAFP Statement Quality Score	0%	0%	44.45%	0%
--	----	----	--------	----

Note. Traditional certification = the educator obtained a four-year teaching degree in the content area being taught and passed all state-required examinations. Alternative certification means the educator obtained certification in the content without obtaining a four-year degree in the content area being taught and meets all state requirements for alternative certification.

Dependent and Independent Variables

The dependent variable is IEP PLAAFP quality, which was determined using the SERC IEP Quality Rubric (2023). The rubric designed by SERC includes all components of IEPs, so researchers isolated and reviewed the PLAAFP quality indicators to ensure compliance with state regulations. The scoring guide for PLAAFP statements included three quality indicators detailed in Table 2. Each PLAAFP statement was graded against each quality indicator, assigned points, and scored as either (0) unacceptable, (1) emerging, (2) progressing, or (3) promising practice. Each PLAAFP statement could earn up to nine points, three possible points per quality indicator. The percentage of quality indicators met was calculated by dividing the total number of points earned by the total number of possible points, then multiplying the result by 100. The State Education Resource Center (SERC, 2013) notes that, before the systematic use of the rubric, all scorers should undergo training, if they have not already done so, on the rubric's contents to support the instrument's reliability. It is also recommended that scorers practice using the rubric and calculate inter-rater reliability (IRR) to assess accuracy and fidelity before implementing systematic use (SERC, 2013). Table 3 provides an example of the PLAAFP statement rubric for indicator 2.

Table 2

Quality Indicators for PLAAFP Statements

Quality Indicator	Description
Indicator 1	“The IEP uses comprehensive general education-driven assessments and benchmarks to isolate and target specially designed instruction to address missing concepts, skills, or strategies that assist students in making progress in general education.”
Indicator 2	“The IEP contains explicit statements determining the student’s strengths as related to the student’s progress in the general education curriculum.”
Indicator 3	“The IEP has an explicit statement of how the dynamic between manifestation of the student’s disability and general education curriculum, instruction, or environment affects access to, participation in, and progress in the general education curriculum.”

Note. Content obtained from “IEP Rubric” by SERC. (2013). ©State Education Research Center.

Table 3

Quality Indicator 2 Scoring Rubric

Scoring	Description
Unacceptable (0)	“There are no or vague, superficial statements about the student’s strengths.” “There are no statements about how the student can participate in the general education settings or curriculum.” “The student’s strengths have no relationship to the general education curriculum, instruction, or environment and are comparative to peers.”
Emerging (1)	“There are general statements about the student’s strengths related to academic learning.” “There are vague statements about the student’s strengths and how the student can participate in general education settings.” “The student’s strengths have no relationship to general education curriculum, instruction, or environment and are not comparative to peers.”
Progressing (2)	“There are general statements about the student’s capacity and strengths to participate and progress in general education curriculum.” “There are general statements about how the student can be successful in general education curriculum.” “The student’s strengths are relative to how he or she can participate in general education curriculum and are not comparative to peers.”

Promising Practice (3)	“There are explicit statements about the student’s capacity and strengths to participate and progress in general education curriculum.” “There are statements about the student’s strengths that can be useful to a general education teacher in understanding how the student can progress in general education curriculum.” “The student’s strengths are relative to how he or she can participate in general education curriculum and are not comparative to peers.”
------------------------	--

Note. Content obtained from “IEP Rubric” by SERC. (2013). ©State Education Research Center.

The independent variable was the PD opportunity, which followed the Behavioral Skills Training (BST) model and was developed by the researcher. The intervention phase included an online training video containing the instruction and modeling phases, weekly in-person coaching sessions, and written feedback on continued rehearsal sessions. Figure 1 provides an example of the written feedback and coaching session form.

Materials

Materials included the recruitment email, consent document, Qualtrics survey for determining participant interest and participation, Excel spreadsheets for scoring PLAAFP statements and data logging and analysis, BST feedback form, online training video, and a Qualtrics social validity survey. The recruitment email, Qualtrics survey for participant interest and participation determination, Qualtrics survey for PLAAFP statement submission, training video, and Qualtrics social validity survey will be emailed to participants in the following order. The BST feedback form was printed and written for each coaching session. The researcher retained a scanned version, and the paper copy will be provided to the participant. All data collected, including quality scores, demographic information for the setting and participants, and scanned copies of feedback forms, were stored on a secure server. During the transition, the printed and completed feedback forms were stored in a secure file container with the researcher. When not in use, they were kept in a locked file cabinet that was inaccessible to anyone outside the research team.

Design

An independent, nonconcurrent, multiple-baseline design across participants was conducted to determine the presence of a functional relation between the BST coaching model and the quality improvement of PLAAFP statements written by special educators in high-need elementary schools. The design consisted of three phases: baseline, instruction, and maintenance. The intervention phase included all four steps of the BST coaching model, while during the maintenance phase, no intervention components were provided to the participants.

Procedures

Recruitment and Participation Determination

Potential participants were selected based on their active employment as special educators at one of the two school locations. The potential participants were asked to complete a short task through a Qualtrics survey. The Qualtrics survey included consent to participate in the study, demographic information, and a prompt to email a de-identified PLAAFP statement they recently wrote as a part of their routine job duties to a designated school district employee, who then reviewed it to ensure FERPA compliance (verify that all names, affiliations, and/or other potentially identifying information redacted) and then provided it to the researchers. Each submitted PLAAFP statement was scored using the designated rubric (SERC, 2013), and a percentage was determined. The first author, a special education administrator and doctoral student in special education, scored each submitted PLAAFP statement. To ensure accuracy, 30% of all submitted PLAAFP statements were double-scored by a research assistant, who underwent an hour-long training with the first author on the contents of the rubric and scoring procedures. An Excel spreadsheet with the quality indicators and scoring procedures was used during the scoring process. Participants scoring 75% or below were eligible for the study. The cut score is based on the current needs of district educators and was determined following conversations with building administrators prior to the implementation of the independent variable. All participants selected to participate in the study were compensated for their time.

Baseline

After consent, baseline data collection began concurrently for all participants. Participants were provided with information about de-identifying material and then prompted to email each PLAAFP statement written during the study to the same designated school district employee. The information regarding the de-identification process consisted of details on what de-identification means, common statement components that require de-identification, and

examples of statements before and after de-identification; see Table 4 for additional details. The process for participants to submit PLAAFP statements throughout the study was the same as the one used during the recruitment process. This process consisted of the participant de-identifying each PLAAFP statement submitted, emailing it to a designated school employee who verified that it was de-identified, and then providing it to the researchers for scoring.

The timeline for data submission varied for each participant, as the number and timing of submissions depended on the participant's individual IEP caseload. During the study, each PLAAFP statement written by the participant as part of their normal job duties was de-identified and submitted to the researchers. The researcher then scored each submitted PLAAFP statement using the designated rubric (SERC, 2013), and a quality percentage was determined. Each quality percentage was logged as a single data point in an Excel spreadsheet and graphed for visual analysis. The participant order was determined based on within-phase stability (Ledford & Gast, 2018). The stability criterion was satisfied, as 80-90% of the data points fell within a 15% range of the average of the data points collected for the phase (Lobo et al., 2017). Upon reaching the stability criterion and having at least three data points for the phase, the first participant entered the intervention phase. The criteria for phase change for the additional participants in both settings were a minimum of two data points for the upcoming participant and the participant in the intervention phase preceding them (Ledford & Gast, 2018), as well as meeting the stability criterion (Lobo et al., 2017).

Table 4

De-Identification Procedures and Guidelines

Section Title	Description
Initial Instructions	Please review the de-identification steps and submit a PLAAFP statement that was written by you, by emailing it to School District Employee. To ensure participant confidentiality, please follow these

	guidelines to de-identify your PLAAFP statements before submission:
De-identification	1. Remove all direct identifiers, including: Names of students, parents, teachers, or schools; Addresses; Dates of birth; Telephone numbers; Email addresses; Student ID numbers or other unique identifiers 2. Replace specific dates with general time frames: Instead of "January 15, 2025," use "Winter 2025" or "Q1 2025" 3. Modify age-specific information: Replace exact ages with age ranges (e.g., "10-12 years old" instead of "11 years old") 4. Generalize test scores and percentiles: Use ranges or general descriptors (e.g., "below grade level" or "in the 80-90th percentile") 5. Remove or generalize unique medical conditions or disabilities: Use broader categories (e.g., "learning disability" instead of a specific diagnosis)
Replacing Information	When removing identifiable information, replace it with generic terms or brackets. Use Student instead of the student's name; Replace teacher names with Teacher or Mr./Ms.
State Assessment Example	Original: "John Doe, a 12-year-old 6th grader at Lincoln Elementary, reads at a 4th-grade level according to the May 2025 assessment." De-identified: "Student, a middle school student, reads below grade level according to the state assessment, taken last spring."
Curriculum Based Measurement Example	Original: "John Doe is currently reading 54 words correct per minute (wpm) on a 2nd-grade reading probe, which is slightly higher than the 2nd-grade fall benchmark of 50 wpm. The 4th-grade fall benchmark is 95 wpm on a 4th-grade reading probe." De-identified: "Student is currently reading 54 words correct per minute (wpm) on a 2nd-grade reading probe, which is slightly higher than the 2nd-grade fall benchmark of 50 wpm. The 4th-grade fall benchmark is 95 wpm on a 4th-grade reading probe."

Intervention Phase

Each participant independently entered the intervention phase and received the online training video through a link to a Canvas page. Completion of the video training was monitored through Canvas page statistics, which showed the amount of time each participant spent

watching the video and engaging with the provided content. Refer to Table 5 for a summary of the online training video's contents. Special education professionals reviewed the content of the training video to ensure its accuracy. The expert panel consisted of university-level professors who taught classes within the special education department. The expert panel provided the lead researcher with comments and suggestions, which the lead researcher completed before using the video and slides with the participants.

The instruction phase, which consisted of a 15–20-minute portion of the video, reviewed the importance and purpose of PLAAFP statements, as well as the required components of a high-quality PLAAFP statement. The required components are general strengths and needs; special considerations including medical, sociocultural, and environmental variables; information from outside agencies such as doctors, therapists, and parents; assessment data that states the student's present level of performance; student-specific factors such as cognitive, behavioral, or adaptive abilities; and a rationale for how the previous information supports the choice in annual goals, services, and supports (Harmon et al., 2020). The modeling phase provided the participants with a video of the researcher modeling the process of reviewing a case study and creating a high-quality PLAAFP.

Table 5

Training Video Contents and Structure

BST Model Steps	Description
Instruction	Participants will watch a video that includes information regarding the importance and purpose of PLAAFP statements and the required components of a PLAAFP statement. The instruction portion of the video will be approximately 10-15 minutes long.
Modeling	Participants will continue with the training video in which the researcher models the steps necessary to create a quality PLAAFP statement. The video will show the process of reviewing collected data to determine strengths and deficits and how to take the

information and create a quality PLAAFP statement. The modeling portion of the video will be approximately 5-10 minutes long.

After watching the training video, each participant in the intervention phase continued to submit written PLAAFPs using the same procedures employed in the baseline phase and engaged in weekly in-person coaching sessions. The researchers scored each submitted statement using the SERC rubric (2013) and determined a quality percentage, which was logged as a data point in an Excel spreadsheet. In addition to submitting written PLAAFP statements, participants engaged in weekly coaching sessions with the first author. During the coaching sessions, the participant and first author reviewed previously submitted PLAAFP statements and discussed the changes needed to meet the quality and compliance standards outlined in the rubric (SERC, 2023). All coaching sessions lasted 15-20 minutes and were conducted in person with the participant and first author at the participant's school location. Sessions were conducted during the participants' designated plan time and were scheduled in advance upon the participant's transition to the intervention phase.

Each coaching session followed the same agenda: (1) review of feedback form, (2) modeling of suggested changes, and (3) participant questions. Participants and the first author signed and dated the feedback form at the end of each session. The feedback form contained the following information: (a) date and time of the coaching session, (b) feedback on previously submitted and scored PLAAFP statements, (c) date and time of the next coaching session, and (f) date and signature lines for the participant and researcher. Figure 1 provides an example of the feedback form and context for the timeline of each coaching session. The fidelity of coaching sessions was conducted for 33% ($n=4$) of the total sessions for each setting (Ledford & Gast, 2018), with a minimum of one session per participant. Fidelity sessions were conducted by trained research assistants and completed via video calls. Implementation fidelity was 100% for

all sessions observed by a research assistant. The criteria for phase change between the intervention and maintenance phases were similar to those for the change between baseline and the intervention phase, with the addition of a clear change in trend. The trend was determined using the split middle method (White & Harring, 1980) to estimate the trend within phases and compare the trends between phases (Ledford & Gast, 2018).

Maintenance Phase

Each participant independently entered the maintenance phase after meeting the criteria for phase change from the intervention phase. During the maintenance phase, participants continued to submit each PLAAFP statement they had written, based on their special education caseload and due dates, following the same procedures used in all other phases. Participants received no interventions during the maintenance phase, unlike those in the intervention phase. The participants did not receive feedback about the PLAAFP statements submitted during the maintenance phase. Participants continued submitting each PLAAFP statement written during the maintenance phase until all participants at the setting in which they were located had completed a minimum of two data points for the maintenance phase or until they no longer had any PLAAFP statements to submit. Upon the last participant entering the maintenance phase and having a minimum of two data points, all participants ceased submitting PLAAFP statements, and data collection was concluded. Upon the conclusion of data collection for all participants in each setting, each participant was briefed on their individual data collected, and their consent for data use was reviewed and verbally reaffirmed or revoked. If consent is revoked, written consent for revocation was gained.

Results

To address the question, is there a functional relation between the use of PD, incorporating a BST coaching model, and increased quality of PLAAFP statements written by special education teachers in high-need elementary schools?, visual analysis and average mean differences suggest that the BST coaching model has a functional relation to the increased quality of PLAAFP statements. The participants' baseline results showed very little variation. Three of the participants had 0% quality scores for all baseline data, and one participant had baseline data ranging from 33.33% to 44.45%. All participants experienced an immediate increase in quality scores upon entering the intervention phase. A total of 57 PLAAFPs were submitted across all three phases and participants. Participant 1 submitted four statements during the baseline, four statements during the intervention, and two statements during the maintenance phase. Participant 2 submitted six statements during the baseline phase, four during the intervention phase, and three during the maintenance phase. Participant 3 submitted eight statements during the baseline phase, four during the intervention phase, and five during the maintenance phase. Participant 4 submitted ten statements during the baseline phase, five during the intervention phase, and two during the maintenance phase. Table 6 details the mean quality score for each participant across all study phases.

Table 6

Mean Quality Score Per Participant Per Phase

Phase	Participant 1 School 1	Participant 2 School 1	Participant 3 School 1	Participant 4 School 2
Baseline	0%	0%	37.5%	0%
Intervention	69.44%	72.22%	86.11%	68.88%
Maintenance	100%	100%	100%	100%

During the intervention phase, the quality percentages increased for each participant compared to their baseline scores. The number of coaching sessions each participant attended varied depending on their individual learning needs, but no participant engaged in more than one coaching session per week. Participant 1 submitted a total of four PLAAFPs during the intervention phase, with quality scores ranging from 33.33% to 100%, and attended four coaching sessions. Participant 2 submitted a total of four PLAAFPs during the intervention phase, with quality scores ranging from 33.33% to 88.89%, and attended two coaching sessions. Participant 3 submitted a total of four PLAAFPs during the intervention phase, with quality scores ranging from 55.55% to 100%, and attended two coaching sessions. Participant 4 submitted a total of five PLAAFPs during the intervention phase, with quality scores ranging from 33.33% to 100%, and attended three coaching sessions. For the maintenance phase, all PLAAFP submissions by all participants received a quality score of 100%. The interquartile range was calculated for all phases and participants to determine variability and identify possible outliers. The results showed little to no variability in the baseline and maintenance phases for all participants. There was a higher level of variability in the scores for the intervention phase among all participants; however, the results show a range of scores with low variability as the scores continued to increase throughout the phase. See Figures 2 and 3 for a graphical representation of the results.

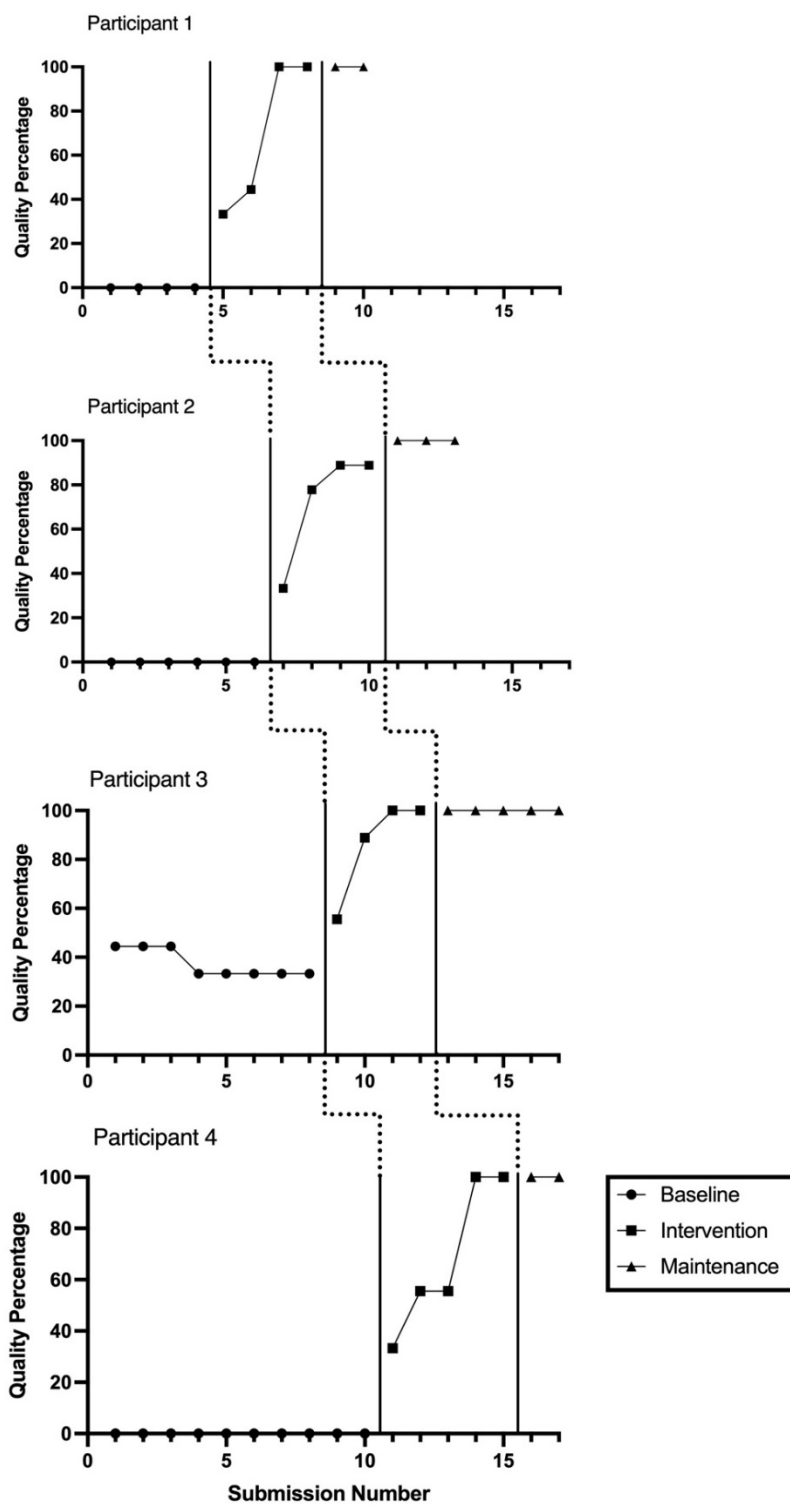
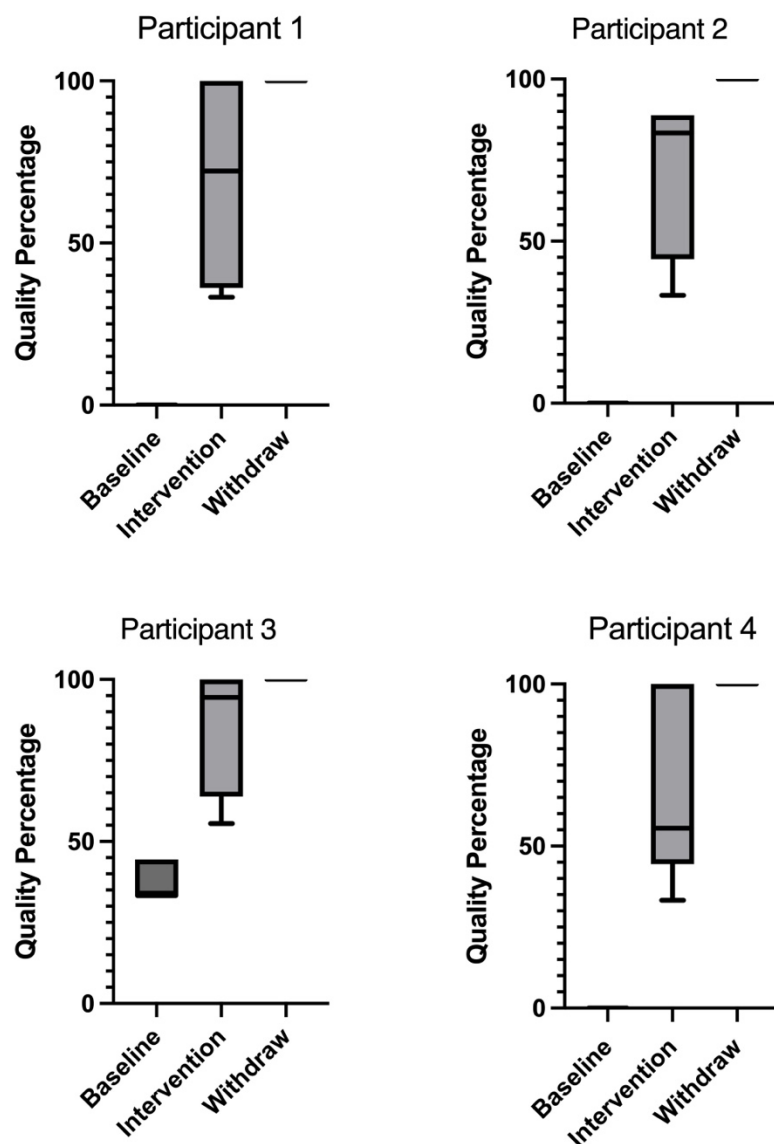
Figure 2*Participant Data*

Figure 3*Interquartile Range*

Note. Figure 3 illustrates the interquartile range for each phase and participant. Data was used to determine variability, phase changes, and participant order.

A visual analysis of the results was used to determine the effects of the intervention, which was scored against the multiple-baseline protocol design proposed by Wolfe and associates (2019). The protocol reviews seven different areas: stability, phase contrast,

immediacy, overlap, consistency, vertical analysis, and staggered treatment. Once the protocol has been completed through dichotomous responses, it provides the researcher with a score for the graphical representation of the data, where 0 indicates no functional relation and 5 indicates a functional relation with strong behavioral change. After completing the protocol for the final visual analysis, the rating scale score was 5, indicating the presence of at least three basic effects, and the data demonstrate a functional relation with large behavioral change (Wolfe et al., 2019). Refer to Table 7 for the completed visual analysis protocol and results.

Table 7

Completed and Scored Four-Tiered Visual Analysis Protocol

Question	Response	Point Value
TIER 1 BASELINE STABILITY: Are there at least 3 data points in baseline for Tier 1 and can you predict the future pattern of the behavior?	Yes	NA
TIER 1 TREATMENT STABILITY: Are there at least 3 data points in treatment for Tier 1 and can you predict the future pattern of the behavior?	Yes	NA
TIER 1 PHASE CONTRAST: Project the trend of Tier 1 baseline into the Tier 1 treatment phase. Is the level, trend, or variability of the data in the Tier 1 treatment phase different than what you would predict based on the baseline data for Tier 1?	Yes	0.75
TIER 1 IMMEDIACY: Is there an immediate change from the last 3-5 data points in baseline to the first 3-5 datapoints in treatment for Tier 1?	Yes	0.18
TIER 1 OVERLAP: Do less than 30% of the treatment datapoints overlap with the baseline datapoints for Tier 1?	Yes	0.18
TIER 2 BASELINE STABILITY: Are there at least 3 data points in baseline for Tier 2 and can you predict the future pattern of the behavior?	Yes	NA
TIER 2 VERTICAL ANALYSIS: Did the baseline data path of Tier 2 change when the treatment began with Tier 1?	No	0
TIER 2 STAGGERED TREATMENT: Was the treatment introduced to Tier 2 AFTER it was introduced to Tier 1?	Yes	NA
TIER 2 TREATMENT STABILITY: Are there at least 3 data points in treatment for Tier 2 and can you predict the future pattern of the behavior?	Yes	NA

TIER 2 PHASE CONTRAST: Project the trend of Tier 2 baseline into the treatment phase. Is the level, trend, or variability of the data in the Tier 2 treatment phase different than what you would predict based on the baseline data for Tier 2?	Yes	0.75
TIER 2 IMMEDIACY: Is there an immediate change from the last 3-5 data points in baseline to the first 3-5 datapoints in treatment for Tier 2?	Yes	0.18
TIER 2 OVERLAP: Do less than 30% of the treatment datapoints overlap with the baseline datapoints for Tier 2?	Yes	0.18
CONSISTENCY: Are the data patterns of treatment for Tier 1 and Tier 2 similar in level, trend, or variability?	Yes	0.18
TIER 3 BASELINE STABILITY: Are there at least 3 data points in baseline for Tier 3 and can you predict the future pattern of the behavior?	Yes	NA
TIER 3 VERTICAL ANALYSIS: Did the baseline data path of Tier 3 change when the treatment began with Tier 1 or Tier 2?	No	0
TIER 3 STAGGERED TREATMENT: Was the treatment introduced to Tier 3 AFTER it was introduced to Tier 2?	Yes	NA
TIER 3 TREATMENT STABILITY: Are there at least 3 data points in treatment for Tier 3 and can you predict the future pattern of the behavior?	Yes	NA
TIER 3 PHASE CONTRAST: Project the trend of Tier 3 baseline into the Tier 3 treatment phase. Is the level, trend, or variability of the data in the Tier 3 treatment phase different than what you would predict based on the baseline data for Tier 3?	Yes	0.75
TIER 3 IMMEDIACY: Is there an immediate change from the last 3-5 data points in baseline to the first 3-5 datapoints in treatment for Tier 3?	Yes	0.18
TIER 3 OVERLAP: Do less than 30% of the treatment datapoints overlap with the baseline datapoints for Tier 3?	Yes	0.18
CONSISTENCY: Is the data pattern of treatment for Tier 3 similar in level, trend, or variability to the data pattern in Tiers 1 and 2?	Yes	0.18
TIER 4 BASELINE STABILITY: Are there at least 3 data points in baseline for Tier 4 and can you predict the future pattern of the behavior?	Yes	NA
TIER 4 VERTICAL ANALYSIS: Did the baseline data path of Tier 4 change when the treatment began with Tier 1, 2, or 3?	No	0
TIER 4 STAGGERED TREATMENT: Was the treatment introduced to Tier 4 AFTER it was introduced to Tier 3?	Yes	NA
TIER 4 TREATMENT STABILITY: Are there at least 3 data points in treatment for Tier 4 and can you predict the future pattern of the behavior?	Yes	NA
TIER 4 PHASE CONTRAST: Project the trend of Tier 4 baseline into the Tier 4 treatment phase. Is the level, trend, or variability of	Yes	0.75

the data in the Tier 4 treatment phase different than what you would predict based on the baseline data for Tier 4?		
TIER 4 IMMEDIACY: Is there an immediate change from the last 3-5 data points in baseline to the first 3-5 datapoints in treatment for Tier 4?	Yes	0.18
TIER 4 OVERLAP: Do less than 30% of the treatment datapoints overlap with the baseline datapoints for Tier 4?	Yes	0.18
CONSISTENCY: Is the data pattern of treatment for Tier 4 similar in level, trend, or variability to the data pattern in at least two previous tiers?	Yes	0.18
Final Score		5

Note: Content obtained from Wolfe, K., Barton, E. E., & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. Behavior Analysis in Practice, 12(2), 491–502. <https://doi.org/10.1007/s40617-019-00336-7>

Descriptive statistics for each PLAAFP submission were also collected: special education category, least restrictive environment (LRE) percentage, and IEP type. For the special education category, 64.9% ($n = 37$) were served under the category of Autism, 22.8% ($n = 13$) were served under the category of Other Health Impairment (OHI), and 12.3% ($n = 7$) were served under the category of Specific Learning Disability (SLD). Participants 1 and 2 both taught in a special education resource room, where students in grades 3-6 received direct instruction services in both academic and adaptive areas. Of the submitted PLAAFP statements for participant 1, 90% ($n = 9$) were subsequent IEPs, with 10% ($n = 1$) being due to an out-of-state move-in, and 80% ($n=8$) were in the general education setting for 40% to 79% and 20% ($n=2$) attending the general education setting for less than 40% of the school day. Participant 1's submissions consisted of 30% Autism ($n=3$), 30% Other Health Impairment ($n=3$), and 40% Specific-Learning Disability ($n = 4$). Participant 2's submissions were 92.3% ($n=12$) subsequent IEPs with 7.7% ($n=1$) being an in-state move-in, and 23.1% ($n=3$) were in the general education setting for more than 80% of their school day, 69.2% ($n=9$) being in the general education setting for 40% to 79% of the day,

and 7.7% ($n=1$) being in the general education setting for less than 40% of their school day.

Participant 2's submissions included 30.8% ($n=4$) Autism, 46.2% ($n=9$) Other Health Impairment (OHI), and 23.1% ($n=3$) Specific Learning Disability (SLD).

Participants 3 and 4 both taught in self-contained special education classrooms for students in kindergarten through sixth grade, where the students engaged in an alternative, adapted curriculum to support their complex learning needs. For participant 3, the submitted PLAAFP statements 100% ($n=17$) were subsequent IEPs and 88.2% ($n=15$) were in the general education setting for less than 40% of their day, and 11.8% ($n=2$) were in the general education setting for 40% to 79% of their school day. Participant 3's submission included 94.1% ($n=16$) Autism and 5.9% ($n=1$) was Other Health Impairment. Participant 4's submissions consisted of 94.1% ($n=16$) subsequent IEPs and 5.9% ($n=1$) were an in-state move-in IEP, while 82.3% ($n=14$) were Autism and 7.7% ($n=3$) were Other Health Impairment. Participant 3's submissions had 82.3% ($n=14$) receiving less than 40% of their daily education minutes in the general education classroom, 11.8% ($n=2$) spending 40% to 70% of their school day in the general education setting, and 5.9% ($n=1$) receiving 80% or more of their daily instruction in the general education classroom. Although participant 4's teaching assignment was to a self-contained life skills classroom, due to the building not being fully staffed for special education teachers, several of the students on her caseload received special education services or supports in a resource class with her as the teacher of record, therefore she was tasked with writing IEPs for students that she did not service.

Interrater Reliability

Each research assistant was trained in procedural safeguards, evidence-based methods for writing PLAAFP statements, and the materials created for the participants. Each research

assistant also received an hour-long training on using the IEP quality rubric to score submissions. The training included rehearsal and feedback components to ensure the research assistants could accurately use the rubric to score participant submissions. Each research assistant received a sample of PLAAFP statements and scored them using the rubric. The first author reviewed the scored PLAAFP statements and filled out rubrics. Any misconceptions within the scoring were discussed and corrected before research assistants used the rubric to score participant data.

A total of 33% ($n=19$) of the total submitted PLAAFP statements across participants and settings were double-scored using the quality rubric, and 30% of each participant's submitted PLAAFP statements were double-scored (Kratochwill et al., 2023). IRR was calculated using Cohen's Kappa. To calculate Cohen's Kappa and determine the level of IRR for each set of double-scored data, the first author created a contingency table to show the quality scores for each rubric category (Cohen, 1960). Next, the observed agreement (P_o) was calculated by summing the agreements and dividing the result by the number of scorings. The expected agreement by chance (P_e) was calculated by using the following formula: $(\text{Total for Rater 1} * \text{Total for Rater 2}) / (\text{Total number of ratings})^2$. This information was applied to Cohen's Kappa formula, $(P_o - P_e) / (1 - P_e)$. Interpretations were determined based on the following guidelines: 0.81-1.00: almost perfect agreement, 0.61-0.80: substantial agreement, 0.41-0.60: moderate agreement, 0.21-0.40: fair agreement, 0.00-0.20: slight agreement, <0.00: poor agreement (Cohen, 1990). The Cohen's Kappa for all double-scored PLAAFP statements was 0.942, falling in the almost perfect agreement range.

Social Validity

After the data review process for all participants had concluded, each was emailed a Qualtrics social validity survey. The survey was anonymous and asked participants about their

experiences throughout the study and their opinions on the effectiveness and feasibility of the interventions. The survey questions and prompts were designed based on Wolf's (1978) social validity framework, which focuses on the social validity of goals, procedures, and outcomes. The survey employed a Likert-type scale (1932), where participants scored statements based on their level of approval or disapproval. The participants were asked to rate each statement on a scale of 1 (strongly disagree) to 5 (strongly agree). Table 8 presents the social validity statements, which were answered using a Likert scale. Additionally, there were several open-ended and optional questions where the participants could type their opinions and statements into the survey. Content analysis (Schreier et al., 2019), a qualitative data analysis method, was used for all written answers. For the content analysis, the researcher reviewed all written submissions, evaluated patterns, identified the frequency of certain words, phrases, or concepts, and then interpreted the data based on the information gathered (Schreier et al., 2019). Table 9 provides the open-ended questions in the social validity survey.

Table 8

Social Validity Survey Statements

Social Validity Survey Statements
The training video effectively improved the PLAAFP statements I wrote.
The in-person coaching sessions effectively improved the PLAAFP statements I wrote.
The in-person coaching sessions were more effective than the online training in enhancing my skills on writing PLAAFP statements.
The training video was a feasible option that fit well with my schedule.
The in-person coaching sessions were more challenging to accommodate in my daily routine than the training video.
The content presented in the training video was relevant to my IEP development needs.
The in-person coaching sessions provided practical, hands-on learning opportunities.
The training video was more time-efficient than the in-person coaching sessions.

Table 9

Social Validity Survey Questions

Social Validity Survey Questions

Describe the most valuable aspects of the training video. How did these compare to the most valuable aspects of the in-person coaching sessions?

In what ways did the training video fit into your schedule? How did this compare to scheduling the in-person coaching sessions?

How did the content relevance of the training video compare to that of the in-person coaching sessions?

How would you describe the quality and usefulness of feedback received during the in-person coaching sessions?

How would you describe the quality and usefulness of the intervention package (video training and coaching sessions)?

The social validity survey was completed by all four participants and retained no information regarding the participants' identities to maintain anonymity. The results from the social validity survey revealed that participants felt the in-person coaching sessions were valuable due to their individualization, targeted feedback, and ongoing support. The participants also stated that the training video was beneficial, fit into their life schedules, and was a useful resource that could be referenced later. The scores from the Likert Scale portion of the social validity survey are detailed in Table 10. The majority of participants agreed that the in-person coaching sessions were more beneficial than the video training. Although participants noted that the video training was an effective way to impart knowledge, this knowledge was further reinforced during the in-person coaching sessions. As for the open-ended questions within the social validity survey, the consensus of the participants was that the feedback form was helpful, as it provided written feedback that could be referenced at a later date by the participants. Overall, the results of the open-ended questions also indicated that the intervention package, video training, and in-person coaching sessions were high-quality, effective, relevant to the needs of the participants, and provided individualized support that helped grow the participants' skills as special educators.

Table 10

Social Validity Likert Scale Results

Social Validity Survey Statements	Average Score	Minimum Score	Maximum Score
The training video effectively improved the PLAAFP statements I wrote.	3.75	3.00	5.00
The in-person coaching sessions effectively improved the PLAAFP statements I wrote.	5.00	5.00	5.00
The in-person coaching sessions were more effective than the online training in enhancing my skills on writing PLAAFP statements.	4.50	3.00	5.00
The training video was a feasible option that fit well with my schedule.	4.25	3.00	5.00
The in-person coaching sessions were more challenging to accommodate in my daily routine than the training video.	2.25	2.00	3.00
The content presented in the training video was relevant to my IEP development needs.	4.25	3.00	5.00
The in-person coaching sessions provided practical, hands-on learning opportunities.	4.25	3.00	5.00
The training video was more time-efficient than the in-person coaching sessions.	3.00	2.00	5.00

Discussion

A nonconcurrent multiple-baseline design across participants was selected due to the nature of the independent and dependent variables. The number of IEPs written by special educators can vary due to several factors, including the grades taught, student needs, caseload size, and changes in enrollment (National Center for Education Statistics, 2024). Due to ethical considerations, the design ensured that each participant received individualized and consistent training and feedback throughout the study. Difficulties such as the intent to avoid increasing

educators' workloads are significant given the current shortage of special educators (Institute of Education Sciences, 2022). Reports indicate that many special educators are leaving the profession due to increased stress levels in the work environment (Institute of Education Sciences, 2022). With that in mind, the researcher aimed to minimize the additional workload on participants. The study design allowed researchers to conduct the study between two elementary schools (Kratochwill et al., 2013; Ledford & Gast, 2018).

Implications for Practice

The results revealed a functional relation between the intervention package and an increase in quality score for PLAAFP statements written by special educators in high-needs public schools. Each participant's engagement with the training video and coaching sessions varied, with no participant exceeding 65 minutes of additional work time over the six weeks. Two participants engaged in as little as 45 minutes of additional work time. The additional work time consisted of the time spent watching the online training video and the minutes spent in the individual coaching sessions. The social validity results align with this information, as participants stated through the survey that the amount of time spent engaging in the intervention package was relatively insignificant compared to the positive outcomes, making their participation in the study worthwhile. This supports the notion that the intervention package, specifically the BST coaching model, can be utilized by public schools to enhance educators' skills in a wide range of instructional, compliance, and practice areas. With the maintenance phase conducted without additional support, such as coaching sessions or feedback, the participants continued to maintain a high level of quality and compliance, demonstrating that the coaching model utilized does not require long-term implementation for participants to sustain the skills taught.

Limitations and Implications for Future Research

There are several limitations that the author would like to note, such as the time of the school year in which the study was conducted. The study was conducted in the last quarter of the school year, resulting in a significantly larger number of IEPs being written compared to the normal caseload seen throughout the year. The school district in which the participants were employed required that all IEPs due during August and September for the following school year be updated prior to the end of the current school year. This policy was in place to ensure that the IEPs implemented at the beginning of the following school year were written based on data from the end of the previous school year and included the correct services and supports that needed to be implemented as soon as school started in the fall. Due to this, the majority of IEPs written by the participants focused on the transition to the next grade level, potentially increasing the number of IEPs being written, which may have led to participants having less time to work on each IEP.

With the time of the year the study was implemented and the school district's calendar having the school year end roughly two weeks earlier than normal, the study was only able to be implemented for a total of six weeks. The number of IEPs each participant was required to write by their employer made it possible for the study to be implemented within this short time. Still, this timeframe also included a large number of abnormal events, such as state testing, awards assemblies, class and school-wide parties, and other end-of-the-school-year events. This may also have affected the amount of time that the participants were able to dedicate to writing each IEP. In addition to the time of the school year, due to the small sample size and the school administrators' awareness of the study, it is unknown whether the school administrators' awareness could have influenced the participants' behavior. Lastly, the study components, such

as the coaching sessions and video training, were completed solely by the first author, outside of the implementation fidelity and IRR sessions. Due to this, it is unknown if the results were generalizable to other staff who might implement the intervention package. The first author also already had a working relationship with all four participants, so it is unknown if the results would be similar if the person implementing the intervention were unfamiliar with the participants.

It would be beneficial for the study to be replicated at the beginning of the school year. This would allow the study to occur over a longer period, ensuring that the carryover of the skill taught is observed, as the current study was only able to gather an average of three maintenance data points per participant. The study could also be replicated on a slightly larger scale to include participants from different school districts, which would allow for the results to be analyzed to determine if the composition and school culture have any effect on the results yielded. This could also be expanded upon by having multiple individuals implement the intervention and conduct coaching sessions, allowing the results to be analyzed to determine if previous learning histories and work relationships yield different outcomes.

Conclusion

The BST coaching model was utilized to increase the quality of PLAAFP statements written by special educators in high-needs elementary schools. All participants engaged in a 25-minute training video and between 2 and 4 coaching sessions, which lasted a maximum of 15 minutes. The training and coaching model was employed in the study to intervene on PLAAFP quality while minimizing additional work for the participants. The results showed a significant increase in quality scores for all participants and indicated that they maintained high levels of quality even after the intervention was removed. This promotes the use of the BST coaching

model for enhancing educators' skill levels and sustaining these skills, making it a cost- and time-efficient training model.

References

- Bateman, B. D. (2011). Individual education programs for children with disabilities. In J. M. Kauffman & D. P. Hallahan (Eds.), *Handbook of special education* (pp. 91–106). Philadelphia, PA: Taylor & Francis/Routledge.
- Burns, M. K., Lembke, E. S., Duesenberg, M. D., Hopkins, S., Hirt, S., Romero, M. E., Thomas, E., Jo Ann Steinbauer, Del Gaiso, A., & Crooks, S. (2023). The relationship between quality of the present level of academic achievement and functional performance statement and student learning. *Journal of Research in Special Educational Needs*, 23(4). <https://doi.org/10.1111/1471-3802.12606>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- DiGennaro Reed, F. D., Blackman, A. L., Erath, T. G., Brand, D., & Novak, M. D. (2018). Guidelines for using behavioral skills training to provide teacher support. *TEACHING Exceptional Children*, 50(6), 373–380. <https://doi.org/10.1177/0040059918777241>
- Doren, B., Flannery, K. B., Lombardi, A. R., & Kato, M. M. (2012). The impact of professional development and student and teacher characteristics on the quality of postsecondary goals. *Remedial and Special Education*, 34(4), 215–224. <https://doi.org/10.1177/0741932512468037>
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015)
- Fingerhut, J., Xu, X., & Moeyaert, M. (2021). Selecting the proper Tau-U measure for single-case experimental designs: Development and application of a decision flowchart. *Evidence-Based Communication Assessment and Intervention*, 15(3), 1–16. <https://doi.org/10.1080/17489539.2021.1937851>

- Finn, J. E., & Kohler, P. D. (2008). A compliance evaluation of the transition outcomes project. *Career Development for Exceptional Individuals*, 32(1), 17–29.
<https://doi.org/10.1177/0885728808315332>
- Flannery, K. B., Lombardi, A., & Kato, M. M. (2013). The impact of professional development on the quality of the transition components of IEPs. *Career Development and Transition for Exceptional Individuals*, 38(1), 14–24. <https://doi.org/10.1177/2165143413489727>
- Flynn, S. D., & Lo, Y. (2015). Teacher implementation of trial-based functional analysis and differential reinforcement of alternative behavior for students with challenging behavior. *Journal of Behavioral Education*, 25(1), 1–31. <https://doi.org/10.1007/s10864-015-9231-2>
- Hall, L. J., Grundon, G. S., Pope, C., & Romero, A. B. (2009). Training paraprofessionals to use behavioral strategies when educating learners with autism spectrum disorders across environments. *Behavioral Interventions*, 25(1), n/a-n/a. <https://doi.org/10.1002/bin.294>
- Harmon, S., Street, M., Bateman, D., & Yell, M. L. (2020). Developing present levels of academic achievement and functional performance statements for IEPs. *TEACHING Exceptional Children*, 52(5), 320–332. <https://doi.org/10.1177/0040059920914260>
- Hogan, A., Knez, N., & Kahng, S. (2014). Evaluating the use of behavioral skills training to improve school staffs' implementation of behavior intervention plans. *Journal of Behavioral Education*, 24(2), 242–254. <https://doi.org/10.1007/s10864-014-9213-9>
- Hott, B. L., Jones, B. A., Randolph, K. M., Kuntz, E., McKenna, J. W., & Brigham, F. J. (2020). Lessons learned from a descriptive review of rural individualized education programs. *The Journal of Special Education*, 55(3), 002246692097267.
<https://doi.org/10.1177/0022466920972670>

- Hott, B. L., Jones, B. A., Rodriguez, J., Brigham, F. J., Martin, A., & Mirafuentes, M. (2019). Are rural students receiving FAPE? A descriptive review of iep's for students with social, emotional, or behavioral needs. *Behavior Modification*, 45(1), 014544551882510. <https://doi.org/10.1177/0145445518825107>
- Institute of Education Sciences. (2022). *School staffing shortages: Results from the January School Pulse Panel*. Institute of Educational Sciences.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-Case intervention research design standards. *Remedial and Special Education*, 34(1), 26–38. <https://doi.org/10.1177/0741932512452794>
- Kurth, J., & Mastergeorge, A. M. (2009). Individual education plan goals and services for adolescents with autism: Impact of age and educational setting. *The Journal of Special Education*, 44(3), 146–160. <https://doi.org/10.1177/0022466908329825>
- Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology : Applications in special education and behavioral sciences* (3rd ed.). Routledge.
- Lee, J. B., & Cherney, L. R. (2018). Tau-U: A quantitative approach for analysis of single-case experimental data in aphasia. *American Journal of Speech-Language Pathology*, 27(1S), 495–503. https://doi.org/10.1044/2017_ajslp-16-0197
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Lobo, M. A., Moeyaert, M., Baraldi Cunha, A., & Babik, I. (2017). Single-Case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, 41(3), 187–197. <https://doi.org/10.1097/npt.0000000000000187>

- Lowman, J. J. (2015). A comparison of three professional development mechanisms for improving the quality of standards-based IEP objectives. *Communication Disorders Quarterly*, 37(4), 211–224. <https://doi.org/10.1177/1525740115595338>
- Miller, I., Crosland, K. A., & Clark, H. B. “Rusty”. (2014). Behavioral skills training with teachers: Booster training for improved maintenance. *Child & Family Behavior Therapy*, 36(1), 19–32. <https://doi.org/10.1080/07317107.2014.878176>
- National Center for Education Statistics. (2024). Students with Disabilities. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved May 30, 2024, from <https://nces.ed.gov/programs/coe/indicator/cgg>.
- National Center for Education Statistics. (2024). *The condition of education*. Nces.ed.gov. <https://nces.ed.gov/programs/coe>
- Rakap, S. (2023). Chatting with GPT: Enhancing individualized education program goal development for novice special education teachers. *Journal of Special Education Technology*, 39(3). <https://doi.org/10.1177/01626434231211295>
- Ruble, L. A., McGrew, J., Dalrymple, N., & Jung, L. A. (2010). Examining the quality of IEPs for young children with autism. *Journal of Autism and Developmental Disorders*, 40(12), 1459–1470. <https://doi.org/10.1007/s10803-010-1003-1>
- Ruble, L., & McGrew, J. H. (2013). Teacher and child predictors of achieving IEP goals of children with autism. *Journal of Autism and Developmental Disorders*, 43(12), 2748–2763. <https://doi.org/10.1007/s10803-013-1884-x>
- Schreier, M., Stamann, C., Janssen, M., Dahl, T., & Whittal, A. (2019). Qualitative content analysis: Conceptualizations and challenges in research practice—introduction to the

- FQS special issue “qualitative content analysis I.” *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3). <https://doi.org/10.17169/fqs-20.3.3393>
- SERC. (2013). *IEP rubric*. State Education Research Center.
- Shriner, J. G., Carty, S., Rose, C. A., Shogren, K. A., Kim, M.-J., & Trach, J. S. (2013). Effects of using a web-based individualized education program decision-making tutorial. *The Journal of Special Education*, 47(3), 175–185.
<https://doi.org/10.1177/0022466912453940>
- Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools. *Journal of Teacher Education*, 52(2), 94–106. <https://doi.org/10.1177/0022487101052002002>
- U.S. Department of Education, Office of Special Education Programs. (2023). *Digest of education statistics, 2022*. Nces.ed.gov.
https://nces.ed.gov/programs/digest/d22/tables/dt22_204.70.asp
- van Vonderen, A., Didden, R., & Beeking, F. (2012). Effectiveness of instruction and video feedback on staff’s trainer behavior during one-to-one training with children with severe intellectual disabilities. *Research in Developmental Disabilities*, 33(1), 283–290.
<https://doi.org/10.1016/j.ridd.2011.07.040>
- White, O. R., & Haring, N. G. (1980). *Exceptional teaching*. C.E. Merrill Pub. Co.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2008). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, 44(1), 18–28. <https://doi.org/10.1177/0022466908328009>
- Wolfe, K., Barton, E. E., & Meadan, H. (2019). Systematic protocols for the visual analysis of single-case research data. *Behavior Analysis in Practice*, 12(2), 491–502.
<https://doi.org/10.1007/s40617-019-00336-7>

Yell, M. L., Katsiyannis, A., Ennis, R. P., Losinski, M., & Christle, C. A. (2016). Avoiding substantive errors in individualized education program development. *TEACHING Exceptional Children*, 49(1), 31–40. <https://doi.org/10.1177/0040059916662204>

Article 3

Enhancing PLAAFP Quality Through the Behavior Skills Training (BST) Model

Recent studies suggest that robust Present Levels of Academic Achievement and Functional Performance (PLAAFP) statements, a crucial and required component of an Individualized Education Program (IEP), are closely linked to favorable student outcomes in mandated benchmark assessments (Burns et al., 2023). This statement is formulated based on up-to-date data. It outlines how the student's disability affects their education, as well as how the team intends to assist the student in accessing a free and appropriate public education (FAPE). PLAAFP statements are essential for implementing the IEP and enhancing student performance (Srivastava, 2024). In the case of *Kirby v. Cabell County School District*, the district court judge determined that if an "IEP fails to assess the student's present levels of academic achievement and functional performance, then the IEP does not comply with IDEA" (2006). This decision highlights the importance of quality PLAAFP statements, as they are essential for meeting IDEA requirements and ensuring that the IEP accurately reflects the student's current needs. In a more recent ruling in *Andrew F. v. Douglas County School District* (2017), it was noted that a student's IEP should be "reasonably calculated to enable a child to make progress appropriate in light of the child's circumstances."

Unfortunately, descriptive study results suggest that IEPs often fail to meet procedural or substantive requirements (see Burns, 2023; Goodwin et al., 2020; Hott et al., 2020; Kurth et al., 2022), thereby placing students with disabilities at risk of not receiving a free and appropriate public education (FAPE). The professional development needs of educators must be considered when attempting to intervene on IEP quality. One effective option for intervening to support educators in developing high-quality PLAAFP statements is behavior skills training (Flynn &

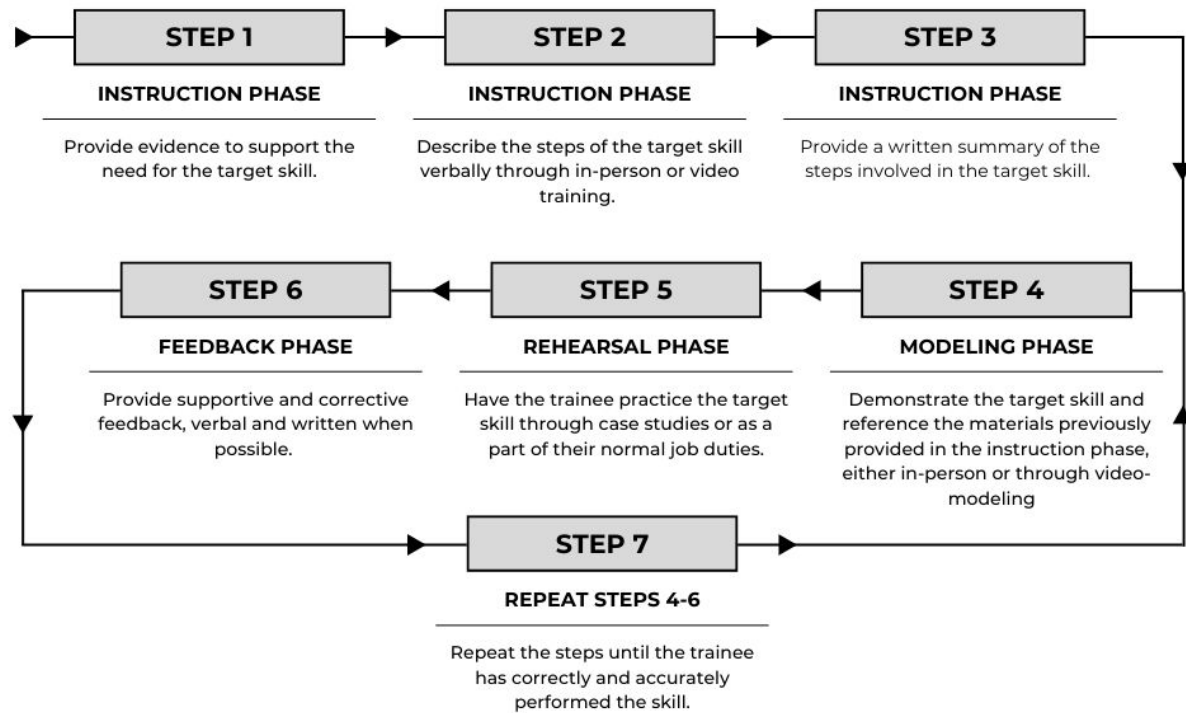
Lo, 2016). This article aims to highlight the importance of high-quality PLAAFP statements, introduce the Behavior Skills Training (BST) model, and provide implementation details on how to use the BST model to increase the quality of PLAAFP statements.

BST

The BST model integrates four phases: (1) instruction, (2) modeling, (3) rehearsal, and (4) feedback (DiGennaro Reed et al., 2018). BST is a systematic approach to coaching and training, providing educators with detailed information about the target skill and opportunities to practice it while receiving expert advice and continuous feedback (Hogan et al., 2014). The BST steps take a cyclical approach where, based on the educator's needs, the steps can be revisited through repetition and refinement (Kirkpatrick et al., 2019). Figure 1 provides a breakdown of each phase of the coaching model. The following sections outline the recommended implementation procedures for crafting high-quality PLAAFP statements using the BST coaching model.

Heather is a special education instructional coach for a large school district where she works with special educators on lesson planning, evidence-based instructional strategies, data collection, and IEP writing processes. More recently, she has received several requests from special educators seeking assistance with writing high-quality PLAAFP statements, following a recent state complaint at a neighboring district. Heather knows that to help each educator truly, she must be organized and systematic in her coaching, so she begins researching different coaching models. She runs across the Behavior Skills Training (BST) model and, after much deliberation, begins designing a training for special educators.

Figure 1. *BST Steps*



Note: Created by the first author using Canva.

Step 1: Instruction Phase- Supporting Evidence

One key element of the BST coaching model is the information provided to the educator during the instruction phase. The first step in the instruction phase is to provide the educator with a rationale for why the training on the target skill is warranted. Although this step might seem minuscule, it is necessary as it provides the educator with the “why,” which has been shown to increase engagement with the training materials and the likelihood of effective implementation (Parsons et al., 2012). When attempting to improve the quality of PLAAFP statements, the rationale can include connections to enhanced student outcomes, as well as the need to comply with federal laws, state-level policies, and procedures. The video training includes all steps within the instruction and modeling phase, and the rationales mentioned are the first component within the training video. The information was incorporated into the first two sections of the training video, starting with an overview of the importance of PLAAFP statements, which highlights the correlation to student outcomes. This is followed by a review of recent legislation that outlines the legal requirements special educators are tasked with meeting.

Since Heather has several special educators who want to participate in the PLAAFP training, she decides that a video training would be the most efficient and effective way to impart knowledge that will lay the groundwork for the remainder of the training steps. Heather ensures that the “why” behind high-quality PLAAFP statements is included, highlighting improved student outcomes and the legal requirements that are at the forefront of all educators' minds, given the recent legal battles in neighboring districts.

Step 2: Instruction Phase- Target Skill

After a rationale has been provided to support the need for training, the next step is to verbally outline the steps required to complete the target skill (Fetherston & Sturmey, 2014).

These steps are commonly referred to as a task analysis, a step-by-step guide for enacting the target skill from start to finish; see Figure 2. It is best practice to review the task analysis verbally with the educator, either in person or through a video training (DiGennaro Reed et al., 2018). The task analysis included all of the required components for the target skill of creating high-quality PLAAFP statements. The required components consisted of (1) student needs, (2) effects on progress in general education, (3) baseline information, (4) connection to goals and services, (5) parent concerns, and, when applicable, (6) transition. The information from steps 1 and 2 was provided to the educators through a video training. Table 1 provides a detailed description of each required component for a PLAAFP statement.

As Heather compiles the information necessary to include in the video training, she wants to ensure that she highlights each required component of a high-quality PLAAFP statement, along with the reasoning that supports the need for each component in relation to positive student outcomes and compliance. In addition to detailing the required components, Heather also includes a task analysis that provides a step-by-step guide for the special educators to follow. For the second step in the BST training model, the instructional phase on the target skill, Heather is ensuring that she verbally states all important information in the video.

Figure 2

Task Analysis for PLAAFP Statement Development

Task Analysis for PLAAFP Statement Development

1. Gather and Review Student Data

- Formal Assessments
- Curriculum-Based Assessments
- Progress Monitoring Data
- Observational Data
- Parent Concerns and Input

2. Identify Key Strengths and Needs (Specific and Objective)

- Academic (e.g., reading, math, writing)
- Functional and Adaptive (e.g., behavior, social/emotional, self-help)
- Related Services (e.g., speech and language, physical therapy)
- Transition (e.g., independent living, post-secondary goals)

3. Impact of Disability

- Describe how the student's disability impacts their ability to access the general education curriculum.
- Describe how the student's disability impacts their ability to be involved and make progress in the general education setting.

4. Data Connection

- Connect the current data to the student's functioning in the educational environment.
- Based on the current data review, identify necessary supports, accommodations, and services
- Determine if adequate progress has been made on previous goals and create new goals.

Table 1

PLAAFP Statement Required Components

Required Component	Description
Student Needs	Include information regarding the student's current academic and functional needs and student strengths. The information should be based on current evaluation and assessment results, student preferences, parent concerns, and state standards.
Effects on Progress in General Education (Disability Impact Statement)	Provide an explanation of how the student's disability affects the student's involvement and progress in the general education curriculum and environment.

Baseline Information	Describe the results of current evaluations and assessments, progress monitoring data, observations, and other forms of data that are appropriate for the student's current developmental stage. The data description should support the identified areas of concern and be correlated to state content standards. Additionally, the data being reviewed should be specific, objective, measurable, and have the ability to be collected frequently to determine whether the student is making progress.
Connection to Goals and Services	The remaining component serves as a bridge between the PLAAFP statement, goals, services, supports, and baseline data. The previous sections should have provided the IEP team with enough information that challenging yet attainable measurable annual goals are created and support the services and supports that are chosen.
Parent Concerns	Parent concerns provide crucial insights in the student's strengths and needs that are unique to them. This information allows the IEP team to tailor the plan, services, and supports to the individual student which promotes positive student outcomes.
Transition (If applicable)	For students who are nearing post-secondary placements, information regarding the student's post-secondary goals and how the IEP team plans to support the student in achieving their goals.

Note: Content adapted from "IRIS | page 6: High-Quality PLAAFP statements. (2020).

Vanderbilt.edu. <https://iris.peabody.vanderbilt.edu/module/iep01/cresource/q3/p06/#content>"

Step 3: Instruction Phase- Written Summary

Once the baseline knowledge for the target skill has been delivered, the next step is to deliver the same information in written form (DiGennaro Reed et al., 2018). The written summary is not only beneficial during the instructional phase, regardless of the method of delivery (e.g., in-person, live video training, pre-recorded video training), but also serves as reference material for educators in later phases and after the training model has commenced. To support educators in writing high-quality PLAAFP statements, the slides used in the training

video, which included task analysis and required component descriptions, were provided in an online, downloadable format.

Heather ensures that all the content she introduces in the training video is available in written format for special educators to use as they go through the training model and for future reference. She also ensures that the written format is legible and easy to understand, focusing on simple fonts and color schemes, as well as providing step-by-step directions in the task analysis (see Figure 2). As an instructional coach, Heather wants to ensure that she meets the needs of all professionals, and she knows that providing information to educators in multiple formats is an evidence-based method to promote knowledge retention in adult learners.

Step 4: Modeling Phase

Once the content has been delivered during the previous steps, the trainer should model the target skill while referencing the previously discussed and provided materials (Hogan et al., 2014). The modeling phase aims to provide a clear and observable example of performing the target skill, which provides clear expectations before the educator practices the skill (Kirkpatrick et al., 2019). The modeling phase can be conducted in person or through video modeling, similar to the previous three steps in the BST process. A case study was used to model the process of training educators on creating high-quality PLAAFP statements. Additional examples outside the modeling case study were used to provide educators with examples of both academic and functional needs. It is essential that during the modeling phase, the trainer demonstrates the target skill in the same context where the educator will apply it.

Additionally, the trainer must ensure that the modeling phase is completed without error and with high integrity. Hence, the educator is once again clear on the expectations and steps needed to perform the target skill (DiGennaro Reed et al., 2018). It is considered best practice for

the trainer to either create a script for the video modeling or conduct the modeling phase in an environment that supports educator understanding, thereby increasing the likelihood that the educator will be able to perform the target skill with minimal errors during the rehearsal phase.

Table 2 provides a summary of best practices for the modeling phase.

Now that Heather has reached the modeling phase, which is incorporated into the recorded training video, she has planned out two modeling examples to use. Since she will be using the training video for a wide range of educators, she has designed the examples to accommodate a multitude of possible scenarios (e.g., elementary-aged students, diverse academic needs, and physical impairments). In the training video, she also ensures that she breaks each example down into the required PLAAFP components to connect the requirements with the outcomes of the example. Heather also color-coded the examples so that educators watching the video training and modeling can correlate each required component with the final written PLAAFP statement used in the modeling example.

Table 2

Summary of Modeling Phase Practices

Practice	Description
Differentiated Delivery Methods	Base delivery method off educator needs and preferences. Modeling can be conducted in person or in-vivo (video modeling), as well as a combination of both methods when repeating the step in later coaching sessions.
Advantages of Video Modeling	The use of video modeling can standardize the training of the target skills, ensuring that all educators are engaged in the same content. It also allows for repeated viewings, can minimize trainer errors through the use of a script, and the video can be easily disseminated to additional educators when needed.
Planning for the Modeling Phase	When planning for the modeling phase it is important to remember to use an appropriate context for the modeling

	example, making it as close to the context the educators will use the target skill in. Additionally, proper planning can ensure that the modeling is error-free and highlights each step needed to complete the target skill.
Educator Engagement	Encourage the educators to ask questions during the modeling phase or after if the modeling phase is done through video. Utilize methods such as verbal praise and positive reinforcement to maintain educator engagement.

Step 5: Rehearsal Phase

The rehearsal phase is essential to bridging the gap between knowledge and performance (Fetherston & Sturmey, 2014). It can be done independently or in conjunction with the previous phases. Additionally, several components can be included in the rehearsal phase that can yield positive results, such as active practice, role-playing, real-life practice, continued guidance and support, repetition, and self-monitoring, which are detailed in Table 3. When intervening on the quality of PLAAFP statements, the rehearsal phase is conducted with real-life practice opportunities and scheduled in advance with the educator. The educators use the skills taught in previous phases to guide their creation of PLAAFP statements written as part of their normal job responsibilities. The repetition of the rehearsal and feedback phases ensured that the skills were retained and generalized to real-world situations (DiGennaro Reed et al., 2018). During the rehearsal phase, whether it's the first session or a repeat, trainers need to have a clear plan or schedule in place. Without a structured approach or consistent timeline for each rehearsal session, inconsistent progress may occur, leading to wasted time. It is important to give educators advanced notice of a scheduled rehearsal session and to agree on the time together. These practices can help build a positive trainer-educator relationship and ensure that the educator feels their time is respected and effectively utilized.

Heather has now sent the training video to each special educator participating in the training model and received confirmation of completion. Now she has begun to schedule weekly coaching sessions where she will support each educator in rehearsing the skills learned with real-life examples. The approach Heather takes with each educator varies depending on the individualized needs. For example, when Heather meets with Mrs. Smith, the rehearsal phase includes more role-playing activities, while Mr. Bradford's rehearsal sessions focus more on self-monitoring. Heather also ensures that she stays organized by maintaining a weekly coaching calendar and holding regular video call office hours, allowing educators to reach out for additional support if needed (see Figure 3).

Table 3

Rehearsal Phase Components

Component	Description
Active Practice	By having educators engage in active practice, where they perform the target skill or behavior in a hands-on manner, the educator can receive guidance and feedback during the rehearsal phase.
Role-Play v. Real-Life	The rehearsal phase can include one or a combination of both role-playing and real-life practice. Role-play can provide the educator with a chance to practice the target skill in a controlled environment with guidance from the trainer. While real-life practice, can provide the educator with a chance to practice the target skill in an environment that is typical and realistic.
Repetition	One important aspect of the BST model is that the educator can engage in the target skill multiple times with the guidance from the trainer. Repetition is important as it provides the educator with the practice needed to build mastery and confidence.
Continue Guidance and Support	The rehearsal phase allows the educator to receive continued guidance, support, and prompts that are specific to their needs as a learner. The trainer has the opportunity to provide

	feedback in a corrective manner or to reinforce the educators performance.
Self-Monitoring	Self-monitoring can be integrated into the rehearsal phase so the educator can have the ability to evaluate their own performance and promote continued learning after the commencement of the formal training model.

Figure 3*Weekly Coaching Schedule Example*

WEEKLY SCHEDULE					
	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
6:00 AM					
7:00 AM			Mrs. Garter Room 34		
8:00 AM					
9:00 AM	Mrs. Smith Room 12				
10:00 AM					
11:00 AM					
12 NOON					Video Call Office Hours
1:00 PM		Mr. Bradford Room 92			
2:00 PM					
3:00 PM					
4:00 PM		Mrs. Jean Video Call			
5:00 PM					
6:00 PM					

Send training video to Mr. Burns → Starting coaching sessions next week.

Step 6: Feedback Phase

The feedback phase provides the educator with specific and constructive feedback on their performance of the target skill. The rehearsal and feedback phases are often conducted in conjunction, providing the educator with immediate feedback that can be actively applied to the practice they are engaging in (DiGennaro Reed et al., 2018). A structured approach to providing feedback, similar to the one seen in the rehearsal phase, can provide the trainer and educator with a method to improve competency in the targeted skill and ensure time is used effectively.

Feedback should be timely, specific, constructive, and tailored to the individual. All feedback should be delivered immediately after the educator practices the skills, or as close to the time the skill was practiced (Kirkpatrick et al., 2019). This can assist the educator in forming a connection between their performance and the feedback provided, while also helping them avoid developing habits that do not align with the task analysis provided in earlier stages of the training. Specific feedback that provides exact details on the wanted and unwanted behaviors involved when the educator practices the skill can also assist in correcting errors and reinforcing correct skills. This also entails that the feedback is constructive, highlighting areas that need improvement while acknowledging strengths, all within a supportive learning environment. Lastly, all feedback must be individualized and tailored to the educator's strengths and needs. Additionally, individualized feedback ensures that it is relevant and actionable for the educator to implement within the next rehearsal phase.

When intervening on the quality of PLAAFP statements, the feedback provided to educators had a delayed component. Still, feedback sessions were held each week to ensure that statements reviewed during the session were written no more than five days before the session. Although best practice is to provide feedback as soon as possible, the skill of writing a high-

quality PLAAFP statement is considered a pivotal behavior, as once it is learned, it can promote continued learning and the generalization of the skill. This means that although the feedback was not immediate, it still coincides with the previous components of the instruction phase, and any feedback provided will continue to build upon the educator's foundation and previous learning experiences to promote continued growth and possible self-management of the skill. In this instance, feedback was provided both verbally in person and in writing through a structured feedback form. Figure 4 provides an example of the feedback form utilized.

During each coaching session, including the rehearsal and feedback phases, Heather completes a feedback form (see Figure 4) for each session. She ensures that she follows the format on the feedback form in each coaching session, so that all educators receive detailed feedback on not only their areas of need but also their areas of strength. Heather also reports the date and time of the next scheduled session in each coaching session, so the educators know when the next session is and what they need to focus on before meeting again. Several of the participating educators highlight that the feedback form is beneficial as it provides them with a reference for later use, especially when Heather is not available to provide in-the-moment feedback.

Step 7: Repeat

The final step in the BST model involves repeating steps 4-6 as needed to best support the educator in acquiring the target skill and meeting the mastery criteria established at the beginning of the training model (DiGennaro Reed et al., 2018). Although this step may seem small, it is crucial to ensure that educators have multiple opportunities to practice the target skill and receive constructive feedback from the trainer. The number of times the steps are repeated is determined by the trainer and based on the educator's needs. During this step, the number of repetitions and the frequency of practice can be adjusted, allowing for additional time between rehearsal and feedback sessions if the trainer and educator agree that more frequent sessions are unnecessary to maintain the skill.

The last step in the BST process is to repeat the steps as necessary for each participating educator. Heather decides to set a criterion for mastery using the task analysis and component descriptions provided in the video training and made available to each educator. During each coaching session, Heather and the educator review previously written PLAAFP statements and work on writing upcoming statements. Heather documents or has the educator document, which includes the required components and steps that were completed. Upon reaching 90% mastery, Heather will gradually reduce her support in naturally occurring situations. For example, Mrs. Smith completed a total of three coaching sessions (repeated steps 4-6 three separate times) before meeting the mastery criteria.

Implementation Strategies

Several additional implementation strategies can be employed throughout the BST model to enhance competency and performance, improve time efficiency, and achieve cost-effectiveness. These strategies can also help mitigate potential challenges that may arise before

and during the implementation of the training model. Professional development (PD) opportunities are typically integrated into the academic calendar of most public schools. Professional development (PD) or continuing education units (CEUs) are required for educators to maintain their certification to teach in a public school setting. One benefit of schools already providing and requiring professional development is that the BST model can be integrated into existing professional development (Saunders et al., 2021). Since the instruction phase typically involves content knowledge and includes examples and steps to complete the target skill, the first phase of the BST model is already complete, and additional steps can be added to enhance educator performance and competency (Shaffer & Thomas-Brown, 2015). Additionally, many traditional professional development (PD) models already include rehearsal and feedback phases and can be easily adapted to better fit the BST model, providing more well-rounded and continuous training.

Establishing a coaching schedule and structured coaching sessions is vital to the BST model (Fetherston & Sturmey, 2014). This component can seamlessly integrate with existing professional development (PD) and continuing education unit (CEU) opportunities, allowing for the continuation of content knowledge delivery through established formats while incorporating coaching components to provide ongoing support to educators. Although including coaching may necessitate additional funding or personnel, these costs can be mitigated through a well-structured coaching schedule (Courtemanche et al., 2020). Furthermore, a structured approach to coaching sessions and other components of the BST model ensures consistency across coaches and within the content and guidance provided to educators. Additionally, technology can facilitate remote coaching sessions. Remote coaching sessions can encompass all steps of the BST model, tailored to the educator's specific needs. This approach also enables coaches or

trainers to conduct sessions at a higher volume by eliminating the need to travel between educator locations (Courtemanche et al., 2020).

Lastly, when utilizing the BST model, trainers or coaches must ensure a systematic method to assess the educator's implementation of the target skill and track improvements or growth, allowing them to measure impact. To assess PLAAFP quality, an evidence-based rubric created by the Connecticut State Department of Education (SERC, 2013) was employed to establish a baseline for all educators and monitor any improvements or potential misconceptions throughout the training model. By using quality rubrics or task analysis to determine if the educator performs the target skill correctly, trainers can leverage this information to guide which steps are repeated, the frequency of coaching sessions, and to ensure that the feedback provided is individualized, specific, and constructive.

Conclusion

The Behavior Skills Training (BST) model can be applied to a wide range of skills and tailored to individual abilities and needs. As mentioned, this training model enhances the quality of PLAAFP statements written by special educators and can be tailored to meet individual needs, utilizing available resources. Although professionals in applied behavior analysis developed the BST model, this systematic approach to training and coaching can be applied in numerous settings, across diverse skills and needs, and with varying time availability and resources.

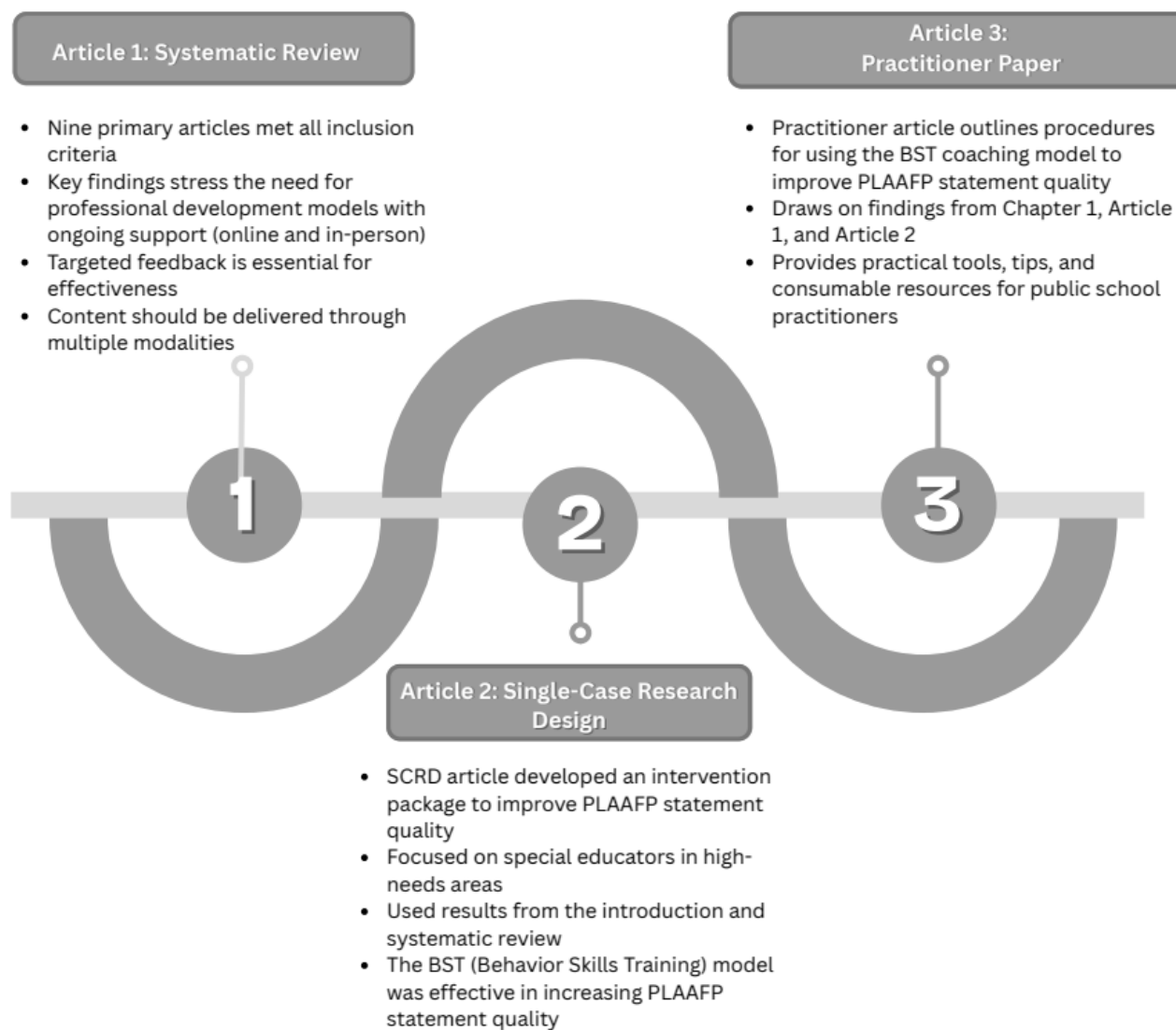
References

- Courtemanche, A. B., Turner, L. B., Molteni, J. D., & Groskreutz, N. C. (2020). Scaling up behavioral skills training: Effectiveness of large-scale and multiskill trainings. *Behavior Analysis in Practice*, 14(1), 36–50. <https://doi.org/10.1007/s40617-020-00480-5>
- DiGennaro Reed, F. D., Blackman, A. L., Erath, T. G., Brand, D., & Novak, M. D. (2018). Guidelines for using behavioral skills training to provide teacher support. *TEACHING Exceptional Children*, 50(6), 373–380. <https://doi.org/10.1177/0040059918777241>
- Fetherston, A. M., & Sturmey, P. (2014). The Effects of Behavioral Skills Training on Instructor and Learner Behavior across Responses and Skill Sets. *Research in Developmental Disabilities*, 35(2), 541–562. <https://doi.org/10.1016/j.ridd.2013.11.006>
- Hogan, A., Knez, N., & Kahng, S. (2014). Evaluating the use of behavioral skills training to improve school staffs' implementation of behavior intervention plans. *Journal of Behavioral Education*, 24(2), 242–254. <https://doi.org/10.1007/s10864-014-9213-9>
- Kirkpatrick, M., Akers, J., & Rivera, G. (2019). Use of behavioral skills training with teachers: A systematic review. *Journal of Behavioral Education*, 28(3), 344–361. <https://doi.org/10.1007/s10864-019-09322-z>
- Saunders, A. F., Wakeman, S., Cerrato, B., & Johnson, H. (2021). Professional development with ongoing coaching: A model for improving educators' implementation of evidence-based practices. *TEACHING Exceptional Children*, 56(6), 004005992110498. <https://doi.org/10.1177/00400599211049821>
- SERC. (2013). *IEP rubric*. State Education Research Center.

Shaffer, L., & Thomas-Brown, K. (2015). Enhancing teacher competency through co-teaching and embedded professional development. *Journal of Education and Training Studies*, 3(3). <https://doi.org/10.11114/jets.v3i3.685>

Discussion

A dearth of Individual Education Program (IEP) plan research suggests a longstanding and pervasive need for educator IEP development training (Finn & Kholer, 2009), IEP implementation (Kurth, 2022), the required components of an IEP (Burns, 2023), and writing standards-based, measurable annual goals (Ruble & McGrew, 2013). The researcher aimed to: (1) synthesize intervention-based studies, (2) test an intervention package, and (3) translate findings for practitioners to explore ways of improving IEP quality and compliance in a time- and cost-efficient manner. Figure 1 provides a summary of how a systematic review, intervention study, and practitioner article built upon one another and guided the subsequent sections.

Figure 1*Dissertation Design Summary*

Note: Created by author using Canva.

Systematic Literature Review

The results of the systematic review revealed few interventions that positively impact IEP compliance and quality. The initial search yielded 8,778 articles and was truncated to those published in or after 2004, as the most recent reauthorization of IDEA occurred in 2004. After removing duplicates and conducting a title and abstract review, a total of 38 articles met the criteria for a full-text review. Upon conducting the full-text review, nine articles met all inclusion criteria to be considered primary articles. All primary articles were coded for study characteristics and methodological rigor. The study characteristics were chosen to provide the researchers with detailed information about the possible replication and feasibility of the interventions. They consisted of the following: (1) year of publication, (2) study authors, (3) funding statements, (4) journal, (5) research design type, (6) independent variable, (7) dependent variable, (8) intervention package details, (9) participant demographics, and (10) locale information.

After reviewing all the primary articles, it was determined that only one article (Burns et al., 2023) met the CEC methodological rigor standards. Additionally, there was a lack of reporting for several methodological rigor standards and study characteristics that are closely related to the ability to generalize the methods and results to other settings and situations. While two articles utilized ChatGPT to evaluate the quality of annual goals (Rakap, 2023; Rakap & Balikci, 2024), the remainder employed some form of training or professional development model to enhance the quality of one or more IEP components (Burns et al., 2023; Doren et al., 2012; Finn & Kohler, 2009; Flannery et al., 2015; Lowman, 2016; Ruble & McGrew, 2013; Shriner et al., 2012).

The main concern with the results of the systematic review was a sufficient description of the training model used as the independent variable to support replication. Although authors provided information regarding the number of workshops, coaching sessions, lecture-based trainings, they did not provide additional details on who conducted the training sessions, their qualifications, or the content provided to the participants. Without this information, the ability to replicate the studies, whether through a research study or in practice, is minimal. The lack of this information also raises concerns regarding the applicability, cost, and time effectiveness of these interventions if they were to be replicated in a more practical setting with the resources available to those attempting to implement the interventions reported.

Despite the aforementioned limitations, several findings provided vital information that was used to develop the PLAAFP development training, using the Behavior Skills Training (BST). For example, when participants continued to access online and in-person support throughout the study, as opposed to just a single training, their quality scores were higher than when educators participated in single trainings (Burns et al., 2023; Ruble & McGrew, 2013; Shriner et al., 2012). With these results in mind, the PLAAFP development training was created to incorporate a training model, Behavior Skills Training, which is adaptable and can include coaching sessions (Doren et al., 2012; Ruble & McGrew, 2013), continued support (Burns et al., 2023; Doren et al., 2012), and different modalities of instruction (Burns et al., 2023), all of which were found to yield increased results.

IEP PLAAFP Development Training

The results of the study revealed that the PLAAFP development training, which followed the BST coaching model and included a minimal time commitment and low cost, enabled special educators to increase the quality of PLAAFP statements written during their normal job duties.

The mean quality score per participant increased with each phase of the study: baseline, intervention, and maintenance. All four participants achieved a 100% quality score for a minimum of three consecutive data points collected during the final phase of the study, maintenance, where participants were not provided with ongoing support and feedback through coaching sessions.

The social validity survey collected information on participants' perceptions of each component of the intervention package, following Wolf's (1978) social validity framework. Participant perceptions of study goals, procedures, and outcomes were captured. The participants unanimously reported that the in-person coaching sessions were effective in improving the quality of PLAAFP statements written by the participants. The participants also reported that the content in the training video and the in-person coaching sessions provided valuable knowledge and feedback that positively affected the quality of the PLAAFP statements written.

The study aimed to enhance the quality of PLAAFP statements written by special educators working in high-needs elementary schools, while ensuring that the intervention was both time- and cost-efficient. Overall, no participants exceeded 65 minutes of additional work outside of their normal job duties over the six-week study period. The maximum amount of time spent by participants in a single week occurred during the first week of the intervention phase, as they were required to watch the 25-minute training video and engage in one 10-minute coaching session. The workload for the intervention agent also changed minimally, as the same training video was provided to each participant. The main additional work time came from preparing for each coaching session by reviewing PLAAFP submissions and outlining the strengths and areas of need to be discussed during the coaching sessions, which estimated on average 10 minutes of additional work per coaching session.

Implications for Researchers

After reviewing the results of both the systematic review and the IEP PLAAFP training, several implications for current and future research should be discussed. Additional research is needed on the quality and compliance of IEPs, as the systematic review yielded only nine articles that attempted to address the quality of IEPs. When such research is done, the study design and final report must focus on being replicable and generalizable to different settings or situations. Without information regarding the intervention agent's qualifications and training, the intervention cannot be replicated, generalized, or expanded upon. Future research should also focus on the time and cost efficiency of the independent variables used to intervene on IEP quality. This knowledge is also vital to the replicability of the intervention, as research that cannot be replicated due to high costs and high time commitments does not benefit the field of special education research and dissemination of empirically validated evidence that supports the use of an intervention in and out of research. Additionally, since one of the main objectives of the dissertation was to enhance the quality of IEPs while also reducing additional work for participants, it would be beneficial to consider frameworks such as Cognitive Load Theory (CLT; Sweller, 2011). CLT highlights the importance of reducing unnecessary or overly difficult load and increasing meaningful load through careful instructional design.

Future research should include additional components in the intervention package, such as self-monitoring through either participants' use of validated quality rubrics or the implementation of detailed task analyses designed to provide a structured list of steps for creating a high-quality IEP as a whole or for a specific IEP component. To further assess the effectiveness of the intervention package, conducting a benefit-cost analysis would provide researchers and readers with a detailed breakdown of the training model's costs compared to its effectiveness. Furthermore, additional empirical research should evaluate not only the quality of

IEP components but also the outcomes associated with high-quality IEPs and the resulting positive academic and behavioral gains for students. Few studies have focused on student academic and social gains. Most of these studies are descriptive reviews, meaning they lack intervention and simply review the quality of IEP components while examining any connections to changes in student outcomes (Kurth & Mastergeorge, 2010; Ruble et al., 2010; Yell et al., 2016). It would be greatly beneficial for educators involved in such research to see directly the positive impacts that high-quality IEPs can have on the students they teach every day.

Implications for Practitioners

Practitioners, including instructional coaches, school and district administrators, educators, and support staff, are in high need of ways to improve the quality of routine tasks such as instructional practices and paperwork compliance. This need has been heightened by the nationwide educator shortage, which has forced schools to hire staff who are not traditionally certified in education, meaning they have not undergone the extensive training that those who received education degrees and maintained teaching credentials (Peyton & Acosta, 2022). Although research has been done to determine the effective components of teacher preparation academies (Leko et al., 2015; Williams, 2016), little has been validated and published on how to reactively support educators who are already in the field and have varying levels of expertise and experience.

The last article in the three-article dissertation is a practitioner article that details the IEP intervention package. The article outlines each step of the BST model and provides guidance on implementing each step as an educational or instructional coach. With the results of the systematic review and IEP PLAAFP training providing supporting evidence for the use of the BST model to coach and train special educators in producing high-quality IEPs, dissemination of the information is vital for expanding the available resources and research on the topic. The

practitioner article also includes information on specific strategies that can be used in conjunction with the BST coaching model such as remote coaching sessions (Courtemanche et al., 2020), structured approaches (Fetherston & Sturmey, 2014), rehearsal methods (DiGennaro Reed et al., 2018), and more detailed information specific to the quality of PLAAFP statements (Srivastava, 2024).

Implications for Policy Makers

Given the state and national progress in education, as well as post-secondary outcomes for students with disabilities (Wilburn & Hoff, 2025), the need for education reform is more pressing than ever. The number of special education enrollments in public schools has increased significantly over the past decade (U.S. Department of Education, Office of Special Education Programs, 2023). Although school enrollments have continued to grow for decades, the rate of increase from year to year has also accelerated, meaning that enrollment naturally rises over time, but at a significantly faster rate in recent years (Arundel, 2025b). In addition to the increase in special education enrollment rates, the nationwide educator shortage has also caused delays in service delivery and evaluation processes, which are both federally mandated and without compliance with the law students are not receiving a free and appropriate public education (FAPE) which affects students with disabilities' ability to make meaningful progress (Arundel, 2025b). Due to educator shortage and subsequent issues, the number of special education due process complaints increased by 16.4% between the 2021-2022 and 2022-2023 school years (Arundel, 2025a). Other forms of special education disputes, such as mediation requests, state complaints, and state-led investigations, were also found to have significantly increased, with a 71.6% increase from the 2012-2013 school year to the 2022-2023 school year nationwide (Arundel, 2025a).

Several factors contribute to the challenges faced by educators, students, and families in public education. The Department of Education's special education budget for the 2025 fiscal year reveals major concerns for all students and educators. The funding provided to public schools by the federal government covers only approximately 10-14% of the true cost of providing a FAPE for students with disabilities. The budget request for the 2025 fiscal year requested \$15.7 billion for special education. Although this number seems substantial, given the current enrollment, the overall budget request equates to approximately \$1,810 per student receiving special education services; an insignificant amount when the cost to educate a single student with a disability can range from \$27,000 to \$82,000 per school year (Lieberman, 2024). This lack of federal funding places a significant financial strain on public schools, as they are tasked with finding ways to continue educating all students, paying school staff a living wage, and maintaining academic and behavioral progress for all students.

Lastly, the recent policy instability, including the proposed dismantling of the US Department of Education and possible conversion of federal IDEA funding into state-level block grants, has begun to threaten access to special education services and support for students with disabilities (Walker, 2025). Changes such as these can lead to inconsistent quality and access, even more so for high-needs areas, and could disproportionately affect students with disabilities. With all of the factors combined, the impact on public education is worrisome. Policy makers are currently tasked with maintaining the legal requirements while also ensuring that the changes they support will lead to positive change in public education, promoting equal access for all students.

Conclusion

The studies included in the dissertation indicate a need for ongoing professional development and support for educators implementing IEPs. Although a wide range of professional development models can be used in school settings, many of them are often costly and require extensive time from educators and trainers. The results of the systematic review highlighted the need for further research to improve the quality of IEPs (see Burns, 2023; Goodwin et al., 2020; Hott et al., 2020; Kurth et al., 2022). The literature reviewed throughout the dissertation also capitalized on the importance of high-quality IEPs, as there is substantial evidence that the quality of an IEP is directly correlated to student outcomes (Burns et al., 2023; Donath et al., 2023).

The findings from the first study were crucial in guiding the design and implementation of the remaining three sections, including the IEP PLAAFP training and the practitioner paper. Overall, the results are promising; however, substantial research is still needed to enhance practice and ultimately improve outcomes for children with disabilities. This work aimed to not only increase IEP quality but to also support improved student outcomes. Advocating for continued support, additional federal funding, and raising awareness of the current state of special education is ever more vital. Policy decisions should be grounded in research and support the educators and children that we serve.

References

- Arundel, K. (2025a, January 27). *Special education formal disputes see notable increases*. K-12 Dive. <https://www.k12dive.com/news/special-education-formal-disputes-school-family-relationships/737830/>
- Arundel, K. (2025b, February 25). *Special education enrollment climbs to nearly 8M*. K-12 Dive. <https://www.k12dive.com/news/number-of-special-education-students-climbs-to-near-8-million/740413/>
- Burns, M. K., Lembke, E. S., Duesenberg, M. D., Hopkins, S., Hirt, S., Romero, M. E., Thomas, E., Jo Ann Steinbauer, Del Gaiso, A., & Crooks, S. (2023). The relationship between quality of the present level of academic achievement and functional performance statement and student learning. *Journal of Research in Special Educational Needs*, 23(4). <https://doi.org/10.1111/1471-3802.12606>
- Courtemanche, A. B., Turner, L. B., Molteni, J. D., & Groskreutz, N. C. (2020). Scaling up behavioral skills training: Effectiveness of large-scale and multiskill trainings. *Behavior Analysis in Practice*, 14(1), 36–50. <https://doi.org/10.1007/s40617-020-00480-5>
- DiGennaro Reed, F. D., Blackman, A. L., Erath, T. G., Brand, D., & Novak, M. D. (2018). Guidelines for using behavioral skills training to provide teacher support. *TEACHING Exceptional Children*, 50(6), 373–380. <https://doi.org/10.1177/0040059918777241>
- Donath, J. L., Lüke, T., Graf, E., Tran, U. S., & Götz, T. (2023). Does Professional Development Effectively Support the Implementation of Inclusive Education? A Meta-Analysis. *Educational Psychology Review*, 35(1). <https://doi.org/10.1007/s10648-023-09752-2>
- Doren, B., Flannery, K. B., Lombardi, A. R., & Kato, M. M. (2012). The impact of professional development and student and teacher characteristics on the quality of postsecondary

goals. *Remedial and Special Education*, 34(4), 215–224.

<https://doi.org/10.1177/0741932512468037>

Finn, J. E., & Kohler, P. D. (2008). A compliance evaluation of the transition outcomes project.

Career Development for Exceptional Individuals, 32(1), 17–29.

<https://doi.org/10.1177/0885728808315332>

Flannery, K. B., Lombardi, A., & Kato, M. M. (2013). The impact of professional development

on the quality of the transition components of IEPs. *Career Development and Transition*

for Exceptional Individuals, 38(1), 14–24. <https://doi.org/10.1177/2165143413489727>

Goodwin, K., Farquharson, K., Yeager Pelatti, C., Schneider-Cline, W., Harvey, J., & Bush, E.

(2020). Examining the quality of individualized education plan (IEP) goals for children with traumatic brain injury (TBI). *Communication Disorders Quarterly*, 43(2),

152574012097611. <https://doi.org/10.1177/1525740120976113>

Hott, B. L., Jones, B. A., Randolph, K. M., Kuntz, E., McKenna, J. W., & Brigham, F. J. (2020).

Lessons learned from a descriptive review of rural individualized education programs.

The Journal of Special Education, 55(3), 002246692097267.

<https://doi.org/10.1177/0022466920972670>

Kurth, J., & Mastergeorge, A. M. (2009a). Individual education plan goals and services for

adolescents with autism: Impact of age and educational setting. *The Journal of Special*

Education, 44(3), 146–160. <https://doi.org/10.1177/0022466908329825>

Kurth, J., & Mastergeorge, A. M. (2009b). Individual education plan goals and services for

adolescents with autism: Impact of age and educational setting. *The Journal of Special*

Education, 44(3), 146–160. <https://doi.org/10.1177/0022466908329825>

- Leko, M. M., Brownell, M. T., Sindelar, P. T., & Kiely, M. T. (2015). Envisioning the future of special education personnel preparation in a standards-based era. *Exceptional Children*, 82(1), 25–43. <https://doi.org/10.1177/0014402915598782>
- Lieberman, M. (2024, August 26). *How much does special education truly cost? Finally, an answer is on the horizon*. Education Week. <https://www.edweek.org/teaching-learning/how-much-does-special-education-truly-cost-finally-an-answer-is-on-the-horizon/2024/08>
- Lowman, J. J. (2015). A comparison of three professional development mechanisms for improving the quality of standards-based IEP objectives. *Communication Disorders Quarterly*, 37(4), 211–224. <https://doi.org/10.1177/1525740115595338>
- Peyton, D., & Acosta, K. (2022). Understanding special education teacher shortages. *State Education Standard*, 22(1), 20–25.
- Rakap, S. (2023). Chatting with GPT: Enhancing individualized education program goal development for novice special education teachers. *Journal of Special Education Technology*, 39(3). <https://doi.org/10.1177/01626434231211295>
- Rakap, S., & Balikci, S. (2024). Enhancing IEP goal development for preschoolers with autism: A preliminary study on ChatGPT integration. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-024-06343-0>
- Ruble, L. A., McGrew, J., Dalrymple, N., & Jung, L. A. (2010). Examining the quality of IEPs for young children with autism. *Journal of Autism and Developmental Disorders*, 40(12), 1459–1470. <https://doi.org/10.1007/s10803-010-1003-1>

- Ruble, L., & McGrew, J. H. (2013). Teacher and child predictors of achieving IEP goals of children with autism. *Journal of Autism and Developmental Disorders*, 43(12), 2748–2763. <https://doi.org/10.1007/s10803-013-1884-x>
- Shriner, J. G., Carty, S., Rose, C. A., Shogren, K. A., Kim, M.-J., & Trach, J. S. (2013). Effects of using a web-based individualized education program decision-making tutorial. *The Journal of Special Education*, 47(3), 175–185. <https://doi.org/10.1177/0022466912453940>
- Srivastava, R. (2024). The influence of individualized education programs (IEPs) on student achievement. *Global International Research Thoughts*, 12(1). <https://doi.org/10.36676/girt.v12.i1.03>
- Sweller, J. (2011). *Cognitive Load Theory*. Springer Science+Business Media, Llc, Cop.
- Walker, T. (2025, February 4). *How dismantling the department of education would harm students* | NEA. Nea.org; National Education Association. <https://www.nea.org/nea-today/all-news-articles/how-dismantling-department-education-would-harm-students>
- Williams, A. R. (2016). Lessons learned from a special education mentor. *Journal of Thought*, 50(12), 49–62. JSTOR. <https://doi.org/10.2307/90010510>
- Yell, M., Katsiyannis, A., Ennis, R. P., & Losinski, M. (2013). Avoiding procedural errors in individualized education program development. *TEACHING Exceptional Children*, 46(1), 56–64. <https://doi.org/10.1177/004005991304600107>

Supplemental Materials

Code	Definition Example
Coder Name	Choose your name
Journal Name	Enter full journal name with no abbreviations. Copy and paste from article <i>Journal of Learning Disabilities</i>
Year Published	Enter the year it was published per the article and citation. Do not enter additional years. <i>2023</i>
Issue Number	Typically the smaller number, inside parentheses <i>43(2)</i>
Volume Number	Typically the large number, before the parentheses <i>43(2)</i>
First Author Last Name	Enter the last name of the first author only <i>Justus</i>
Additional Authors	Enter the last name of the additional authors in the order they appear, separated by a comma <i>Atwood, Martin, Muns</i>
Full Article Title	Enter the full article title with no abbreviations Copy and paste from article <i>The relationship between quality of the present level of academic achievement and functional performance statement and student learning</i>
Did the authors report having funding for the article/study?	Yes: if the authors report funding from private or public sources, such as IES, OSEP, CEC
Funding Statement (Copy and Paste)	<i>The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially supported by the Office of Special Education Programs, U.S. Department of Education [Student Initiated Grant].</i> <i>This work was supported by Grant Number R34MH073071 and RC1MH089760 from the National Institute of Mental Health. The content is solely the responsibility of the</i>

	<i>authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.</i> No: if the authors report no funding was received Leave blank if no
Study Design	Cluster RCT: A cluster randomized trial (CRT) is a randomized controlled trial in which pre-existing groups, called clusters, of individuals are randomly allocated to treatment arms. For example, clusters may be clinical practices or schools where the individuals are patients and school children, respectively. RCT: In an RCT, trainees are randomly assigned to receive 1 of 2 or more educational interventions. Randomized controlled trials are quantitative, comparative, controlled experiments in which treatment effect sizes may be determined with less bias than observational trials. QED: A quasi-experiment is an empirical interventional study used to estimate the causal impact of an intervention on target population without random assignment. Quasi-experimental research shares similarities with the traditional experimental design or randomized controlled trial, but it specifically lacks the element of random assignment to treatment or control. Instead, quasi-experimental designs typically allow the researcher to control the assignment to the treatment condition, but using some criterion other than random assignment (e.g., an eligibility cutoff mark) Pre/Post: A before-and-after study (also called pre-post study) measures outcomes in a group of participants before introducing a product or other intervention, and then again afterwards. Any changes in the outcomes are attributed to the product or intervention. These studies typically lack the components that are included in the others such as randomization and having a control group.
Study Purpose	Copy and paste the purpose statement.

	<i>The purpose of this study is to identify the relative importance of these four domains in understanding variability in educational outcomes, as defined by achievement of IEP goals, of children with autism who were randomized into the experimental condition of one of two randomized controlled studies of a manualized parent–teacher consultation implementation intervention called the Collaborative Model for Promoting Competence and Success (COMPASS; Ruble et al. 2012a).</i>
Did the authors report research questions? Copy and Paste the research questions.	Yes-research questions are clearly stated and reported No-research questions are not reported within the entirety of the manuscript
Independent Variable	The independent variable is the variable that is varied or manipulated by the researcher. In most cases, this is the intervention. <i>Multi-Component Professional Development on Transition Plans</i>
Dependent Variable	Dependent Variable: The variable that depends on other factors that are measured. These variables are expected to change as a result of an experimental manipulation of the independent variable or variables. <i>Goal Quality of Postsecondary Transition Goals</i>
Did the researchers create their own rubric for determining IEP quality?	Yes-Researchers used components from other rubrics to design the rubric used for the study OR researchers used IDEA/State regulations to design used for the study OR a mix of both. No-Researchers clearly state the name of the rubric used and it was not adapted in any way. If it was adapted in any way to be used in the study, choose YES and include the details in the next question.
IEP Quality Rubric (If YES was chosen above)	Enter the components used, creation process, rubrics sampled from, IDEA regulations used, etc. in the making of the rubric used for the study. <i>The first two objectives for each child, totaling six objectives per participant (2 objectives × 3 children),</i>

	were evaluated using the following criteria adapted from Notari-Syverson and Schuster (1995) and the mandates set forth by IDEA (2004): 1. The objective relates to the scope and sequence of state educational standards. 2. The outcome of the behavior is meaningful for improving the child's communication. 3. The objective includes a (a) behavior, (b) condition, and (c) criterion.
IEP Quality Rubric Used (If NO was chosen above)	Copy and Paste <i>R-GORI</i> Enter the following if the rubric was created by the researchers <i>Researcher Created</i>
Intervention Package	Detail the intervention used, note if multiple interventions were used, and provide information to detail the intervention package as a whole. <i>If this is a longer section of the paper, then enter one to two sentences for each component.</i> <i>Delivery of web-based materials. Previously described training content was presented in text form accessed via a password protected website for use in the present study.</i> <i>Delivery of workshop. The WO and PC groups simultaneously attended a 1-day workshop.</i> <i>Morning session: Developing educationally relevant IEP goals and objectives. PowerPoint lecture, group discussion, examples, case studies, presenter modeling, and handouts were used to present the three strands of knowledge and skills previously described.</i>
Number of IEPs reviewed throughout the entirety of the study.	List the number of IEPs that the authors used to determine effectiveness of the intervention and determine baseline.
Parts of the IEP analyzed in the study, based on the quality rubric used.	If multiple components are analyzed, list each one separated by a comma.

	<i>PLAAF, goals, transition plans</i>
Did the authors report the coding procedures for coding IEP quality?	Yes-procedures for coding IEP quality are included No-procedures for coding IEP quality are not included, only results and information about the rubric.
Enter the coding procedures for coding IEP quality.	Yes-Copy and Paste <i>Each criterion, with Criterion 3 having three individual pieces scored separately (behavior, condition, criterion), was scored using a rating of 0, 1, or 2, for a total possible score of 10 per objective. A total of 13 objectives were evaluated per participant—6 prior to training, 1 at post-training, and 6 at post-coaching.</i> <i>Given that the PC group's post-coaching objectives could potentially be influenced by the extra hour of training they received, participants in all three groups submitted a PLEP statement, goal, and objectives for one child on their caseload 1 week following the workshop.</i> No- leave blank
Did the authors report the training procedures for those coding IEP quality?	Yes-training procedures for coding IEP quality are included. No-training procedures for coding IEP quality are not included.
Enter the training procedures for those coding IEP quality.	Yes-Copy and Paste <i>The first and second author coded all IEPs using the quality rubric, independently. A graduate assistant coded 50% of the IEPs and underwent a one hour training where them and the first author reviewed the coding manual.</i> No-leave blank
Did the authors report qualification of those coding IEP quality?	Yes- the qualifications of those coding IEP quality was reported by the authors (highest degree earned, certifications, years of experience) No- the qualifications of those coding IEP quality was not reported by the authors
Enter the qualifications of those coding IEP quality	Yes- Copy and Paste <i>The graduate assistant coding was a certified BCBA and a special educator for 6 years.</i> No-leave blank

Did the authors report IOA/IRR for IEP quality coding?	Yes- The authors report IOA/IRR for the IEP quality coding. No-The authors did not report IOA/IRR for the IEP quality coding OR it was mentioned but not IOA/IRR percentages were reported.
Enter the IOA/IRR for coding IEP quality.	Yes- Copy and Paste <i>IOA was collected for 30% of the IEPs coded. All IOA sessions reached 100% agreement.</i> No-leave blank
Overall results/effects of the intervention.	(Copy and Paste) <i>Analysis of the predictors multivariate revealed a slightly different picture. When all child variables were examined in the regression, only child engagement emerged as a significant independent predictor of outcome. This finding was unexpected because IQ, language, and autism severity have been identified in numerous studies as pretreatment predictors of outcome. This suggests that the impact that these other more typically measured child variables have on outcomes may be produced, at least in part, through their influence on engagement.</i>
Did the authors report locale information for the schools/districts participating?	Yes- Ex. urban, suburban, rural, mid-size city, western state, southern location, etc. No-no locale information
Did the authors identify a country and/or region?	Yes- Ex. Minnesota, south-central united states No-leave blank
What was the location type? (Check all that apply)	Urban Suburban Rural Other: (fill in) Not mentioned <i>If it is not clearly stated, do not infer based on other information, only click the ones that are clearly stated.</i>

Did the authors use a descriptor of the location?	Yes- Ex. Mid-sized city No- location descriptor not included
Copy the location descriptor here.	<i>Mid-sized city</i> No-Leave blank
Did the authors include any additional locale information?	Yes- Ex. <i>Six schools across the same county in Missouri</i> No-no additional locale information was included
Copy additional locale information here.	Yes- <i>Six schools across the same county in Missouri</i> No-leave blank
Demographics	Student Participant Demographics Treatment Group Student Participant Demographics Control Group Teacher Participant Demographics Treatment Group Teacher Participant Demographics Control Group <i>Follow the detailed instructions throughout the sections.</i> <i>If additional demographics were included, click other and enter the title of the demographic and results reported</i>
CEC Quality Standards	Description/Examples/Justification
1.0 Context and Setting	Context and setting. The study provides sufficient information regarding the critical features of the context or setting.
1.1	"1.1. The study describes critical features of the context or setting relevant to the review; for example, type of program or classroom, type of school (e.g., public, private, charter, preschool), curriculum, geographic location, community setting, socioeconomic status, physical layout." For all CEC 2014 quality indicator components: "when reviewers can reasonably infer that the quality indicator is met on the basis of other, related information reported—they can decide that a study has met a quality indicator,

	even if the research report does not explicitly report addressing it" (p. 2). Don't need "how many windows," but the study should at least describe who the people are in the room. "Describing the context and setting of studies is primarily important for evidence-based reviews to determine whether a study falls within the parameters of the review. Therefore, to address QI 1.1, we require only that researchers report the context/setting in sufficient detail to determine whether the study should be included in the review" (p. 223). Critical features of the physical setting are described with sufficient precision to allow replication. "Another researcher should be able to use the description of participants and setting to recruit similar participants who inhabit similar settings" (p. 166-167).
2.0 Participants	2.0. Participants: "The study provides sufficient information to identify the population of participants to which results may be generalized and to determine or confirm whether the participants demonstrated the disability or difficulty of focus."
2.1	"2.1. The study describes participant demographics relevant to the review (e.g., gender, age/grade, race/ethnicity, socioeconomic status, language status)." At least one demographic element needs to be described. "Similar to context and setting, QI 2.1 requires that study authors describe characteristics of participants sufficiently to determine whether the study should be included in a review" (p.223)
2.2	"2.2. The study describes disability or risk status of the participants (e.g., specific learning disability, autism spectrum disorder, behavior problem, at risk for reading failure) and method for determining status (e.g., identified by school using state IDEA criteria, teacher nomination, standardized intelligence test, curriculum-based measurement probes, rating scale)."

"In addition, reviewers should only consider participants who are part of population targeted by their review. For example, for a review targeting learners categorized as having emotional or behavioral disorders (EBD), in an ABAB study involving three participants with EBD and two with learning disabilities, reviewers only consider (i.e., visually analyze) the outcomes for participants with EBD" (p. 8).

Intellectual disability paired with an IQ score is enough, even if the IQ test is not named, because there are very few IQ tests and all are valid and reliable (as apposed to the great variability of reading assessments).

"Diagnosis of..." is not enough.

Also, e.g. items like "teacher nomination" need additional, non-vague details.

"...move beyond school district-provided labels. Informing readers that students met the state criteria for identifying students as developmentally delayed provides little information on the type or extent of disability a student may have. Researchers need to provide a definition of the relevant disability(ies) or difficulties and then include assessment results documenting that the individuals included in the study met the requirements of the definition. ...link their definitions to those in the current literature" (p. 154).

Additional info may include comorbidity, demographics, academic assessments (with ESs), % receiving subsidized lunch. "...group difference on salient variables must also be presented" (p. 155).

"To meet QI 2.2, study authors must describe (a) the disability/risk status of participants and (b) either report how disability/risk status was determined or provide information from which determination of disability/risk status can be inferred. Reporting that participants were identified as having a disability using state or district criteria meets QI 2.2. Reporting that participants have a disability specified by Individuals With Disabilities Education Improvement Act (IDEA) and are served in a special education setting/receiving special education

	services also meets QI 2.2, as it can be reasonably inferred that participants were identified as having the disability using state/district criteria. Studies that describe participants as having non-specified disabilities, having a specific disability (e.g., autism), or being at risk without providing additional information indicating how the disability/risk status was determined do not meet QI 2.2" (p. 225).
3.0 Intervention Agent	3.0. Intervention agent. The study provides sufficient information regarding the critical features of the intervention agent.
3.1	"3.1. The study describes the role of the intervention agent (e.g., teacher, researcher, paraprofessional, parent, volunteer, peer tutor, sibling, technological device/computer) and, as relevant to the review, background variables (e.g., race/ethnicity, educational background/licensure). " One or more background variables (as relevant to the purpose of the review). "Sufficient information describing important characteristics of the intervention providers is supplied, and appropriate procedures to increase the probability that intervention providers were comparable across conditions are used." "...so that readers can understand the type of individual who may be capable of administering the intervention... Relevant characteristics may include age, sex, race, educational background, prior experience with related interventions, professional experience, and number of children with and without disabilities in the family (for parents)" (p. 156). "We require that (a) the role of the interventionist is noted or can be reasonably inferred and (b) interventionists are described so it can be determined whether the study should be included in the review (QI 3.1). ... To meet the latter condition for a review of the effectiveness of a practice for secondary students with learning disabilities delivered by paraprofessionals, reviewers must be able to tell whether interventionists were paraprofessionals. If a review is not specific to a population of interventionists, the latter condition does not apply" (p. 225).

3.2	3.2. The study describes any specific training (e.g., amount of training, training to a criterion) or qualifications (e.g., professional credential) required to implement the intervention, and indicates that the interventionist has achieved them. If the study author is the interventionist and designed the intervention, this QI is met (e.g., self-trained by virtue of designing). "It may also be useful to present information on assessments involving knowledge of the topic to be taught, knowledge of appropriate pedagogical methods, efficacy measures, and/or attitudinal measures" (p. 156). "If specific training or qualifications are required to implement a practice appropriately, they must be noted and interventionists must have achieved them (QI 3.2). Describing training and noting that it was delivered to/ completed by interventionists are acceptable. ...QI 3.2 is still met if (a) study authors note that no training or qualifications were required or (b) reviewers consider the intervention can be reasonably implemented without special training/ qualifications" (p. 226).
4.0 Description of Practice	4.0. Description of practice. The study provides sufficient information regarding the critical features of the practice (intervention), such that the practice is clearly understood and can be reasonably replicated.
4.1	4.1. The study describes detailed intervention procedures (e.g., intervention components, instructional behaviors, critical or active elements, manualized or scripted procedures, dosage) and intervention agents' actions (e.g., prompts, verbalizations, physical behaviors, proximity), or cites one or more accessible sources that provide this information. "4.1. The study describes detailed intervention procedures (e.g., intervention components, instructional behaviors, critical or active elements, manualized or scripted procedures, dosage) and intervention agents' actions (e.g., prompts, verbalizations, physical behaviors, proximity), or cites one or more accessible sources that provide this information."

	"The intervention is clearly described and specified." "Researchers should provide a precise description of the independent variable to allow for systematic replication as well as facilitate coding in research syntheses such as meta-analysis" (p. 156). "Reporting every detail about a practice is neither feasible nor required to meet QI 4.1. For example, a checklist of critical elements including interventionist actions that enables readers to understand how to implement the essential aspects of the intervention satisfies this QI. Merely naming the practice is unacceptable, although providing one or more citations to publicly available sources appropriately describing the practice is acceptable" (p. 226).
4.2	4.2. When relevant, the study describes materials (e.g., manipulatives, worksheets, timers, cues, toys), or cites one or more accessible sources providing this information. "4.2. When relevant, the study describes materials (e.g., manipulatives, worksheets, timers, cues, toys), or cites one or more accessible sources providing this information." "Interventions should be clearly described on a number of salient dimensions, including conceptual underpinnings; detailed instructional procedures; teacher actions and language (e.g., modeling, corrective feedback); use of instructional materials (e.g., task difficulty, example selection); and student behaviors (e.g., what students are required to do and say)" (p. 156). "QI 4.2 is not applicable if no materials are mentioned and none appears to be needed to implement the practice. The QI is met by merely noting materials used if materials are common, self-explanatory, and appear to be used in a typical manner (e.g., a stopwatch used to time duration of behavior does not need to be described further)" (p. 226).
5.0 Implementation Fidelity	The practice is implemented with fidelity.
5.1	5.1. The study assesses and reports implementation fidelity related to adherence using direct, reliable measures (e.g., observations using a checklist of critical elements of practice).

	Direct observation components, TI checklist. If a script is used, must say was followed word for word (or similar use of quantifiable levels). "Fidelity of implementation is described and assessed in terms of surface (the expected intervention is implemented) and quality (how well the intervention is implemented) features." "Fidelity of implementation is described and assessed in terms of surface (the expected intervention is implemented) and quality (how well the intervention is implemented) features. ...Observations should take place over the entire course of the study on a regular basis. The research team should include some measure of interobserver reliability. Key features to look for in assessing this criterion are: (a) inclusion of key features of the practice... ; (b) adequate time allocation per day or week for the intervention; and (c) coverage of specified amount of material in the curriculum or teacher guide (when appropriate). ... More sophisticated measures of fidelity require an observer to document not only the occurrence of prescribed steps in the process, but also that the interventionist employed the techniques with quality" (p. 157). "We require that adherence to implementation procedures be directly and reliably assessed and reported (QI 5.1). Typically, this is accomplished through observations using a checklist of critical elements. Because the field has not agreed on a particular level of acceptable adherence (Harn, Parisi, & Stoolmiller, 2013; Johnson & McMaster, 2013), we do not require a specific level of adherence—only that it be assessed and reported" (p. 226).
5.2	5.2. The study assesses and reports implementation fidelity related to dosage or exposure using direct, reliable measures (e.g., observations or self-report of the duration, frequency, curriculum coverage of implementation). If 5.1 is present, dosage is described somewhere (but not necessarily reported via a checklist), and graph is present, then QI 5.2 is met.

	"Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?" (D) "Fidelity of implementation is described and assessed in terms of surface (the expected intervention is implemented) and quality (how well the intervention is implemented) features. ...Observations should take place over the entire course of the study on a regular basis. The research team should include some measure of interobserver reliability. Key features to look for in assessing this criterion are: (a) inclusion of key features of the practice... ; (b) adequate time allocation per day or week for the intervention; and (c) coverage of specified amount of material in the curriculum or teacher guide (when appropriate). ... More sophisticated measures of fidelity require an observer to document not only the occurrence of prescribed steps in the process, but also that the interventionist employed the techniques with quality" (p. 157). "Similarly, no particular threshold of dosage/exposure must be achieved to address QI 5.2. Rather, dosage or exposure must be assessed and reported, or information from which dosage or exposure can be inferred must be reported. Describing how long intervention sessions lasted or how many days/classes the intervention occurred is acceptable (p.226)
5.3	5.3. As appropriate, the study assesses and reports implementation fidelity (a) regularly throughout implementation of the intervention (e.g., beginning, middle, end of the intervention period), and (b) for each interventionist, each setting, and each participant or other unit of analysis. "5.3. As appropriate, the study assesses and reports implementation fidelity (a) regularly throughout implementation of the intervention (e.g., beginning, middle, end of the intervention period), and (b) for each interventionist, each setting, and each participant or other unit of analysis. If either adherence or dosage is assessed and reported, this item applies to the type of fidelity

	assessed. If neither adherence nor dosage is assessed and reported, this item is not applicable." "Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?" (D) "Researchers should ideally (a) specify when, where, and for whom fidelity is assessed and (b) report implementation separately for each level of these variables. However, indicating that fidelity was assessed across relevant elements and reporting aggregate measures of fidelity (e.g., average percentage of critical elements implemented) meet QI 5.3. For example, noting that adherence and dosage were assessed at the beginning, middle, and end of a study for each interventionist and reporting aggregate measures of fidelity are acceptable for a group comparison study involving multiple interventionists and conducted in one setting" (p. 226).
6.0 Internal Validity	6.0. Internal Validity: "The independent variable is under the control of experimenter. The study describes the services provided in control and comparison conditions and phases. The research design provides sufficient evidence that the independent variable causes change in the dependent variable or variables. Participants stayed with the study, so attrition is not a significant threat to internal validity."
6.1	"6.1. The researcher controls and systematically manipulates the independent variable." Without TI reported, cannot have internal validity. If 5.1 = 0 then 6.1 and 6.7 = 0. "Implementation of the practice must be under the control of the researcher (QI 6.1). Retrospective studies (e.g., natural experiments) do not meet this QI" (p. 226).
6.2	6.2. The study describes baseline (single-subject studies) or control/comparison (group comparison studies) conditions, such as the curriculum, instruction, and interventions (e.g., definition, duration, length, frequency, learner: instructor ratio).

	Key word is <i>describes</i>, so okay if no TI. "The nature of services provided in comparison conditions are described and documented." "Was any documentation of the nature of instruction or series provided in comparison conditions?" "...researchers should also describe, assess, and document implementation in comparison groups. At a minimum, researchers should examine comparison groups to determine what instructional events are occurring, what texts are being used, and what professional development and support is provided to teachers" (p. 158). "Although detailed information on comparison conditions is desirable, reporting every detail is neither feasible nor required to meet QI 6.2. Reporting information that enables readers to understand the essential aspects of the conditions and instruction involved in comparison condition(s) or baseline/comparison phase(s) satisfies this QI. Merely noting that the comparison conditions were "business as usual" is unacceptable" (p. 226).
6.3	6.3. Control/comparison-condition or baseline-condition participants have no or extremely limited access to the treatment intervention TI checklist not needed, but somehow needs to be explicitly stated, especially during withdrawal. "Other factors to assess include possible access to the curriculum/content associated with the experimental group's intervention, time allocated for instruction, and type of grouping used during instruction" (p. 158). "When exposure to the intervention is unlikely for comparison groups or during baseline/comparison phases, QI 6.3 is addressed without reporting any additional information. For example, the QI is met automatically in a group comparison study if the comparison group received instruction in a different school or classroom from the treatment group" (p. 226).

	If pre/post intervention design, mark “other” and enter “Participants served as their own control” Then code as a 1
6.4	"6.4. The study clearly describes assignment to groups, which involves participants (or classrooms, schools, or other unit of analysis) being assigned to groups in one of the following ways: (a) randomly; (b) nonrandomly, but the comparison groups are matched very closely to the intervention group (e.g., matched on prior test scores, demographics, a propensity score; see Song & Herman, 2010); (c) nonrandomly, but techniques are used to measure differences and, if meaningful differences are identified—for example, statistically significant difference, difference greater than 5% of a standard deviation (What Works Clearinghouse, 2011)—to statistically control for any differences between groups on relevant pretest scores or demographic characteristics (e.g., statistically adjust for confounding variable through techniques such as ANCOVA or propensity score analysis); or (d) nonrandomly on the basis of a reasonable cutoff point (regression discontinuity design)." "Although researchers should ideally report the specific procedures used when randomly assigning participants to groups (e.g., online random number generator), simply reporting that units of analysis were randomly assigned is acceptable to address QI 6.4. Functionally random processes (e.g., school administered lottery) meet this QI (WWC, 2011). With the exception of regression discontinuity designs, some indication of functional equivalence of groups is required when groups are non-randomly assigned. Groups may be matched (e.g., on pretest scores, on a propensity score) or differences between groups on relevant variables can be assessed and, if meaningful differences are identified, controlled for statistically. ... Studies using non-random assignment that report groups are equivalent or comparable without providing details to support the claim do not meet QI 6.4" (p. 226-227)
6.8	Overall attrition is low across groups (e.g., < 30% in a 1-year study).

	If the study was not optional (students had no choice to withdraw or not), this QI is NA. "Considering the lack of empirical guidance, we suggest reviewers use a liberal criterion for overall attrition (QI 6.8), such as <30% (Gersten et al., 2005). Reviewers should consider the context of the studies reviewed when setting a criterion for attrition (see Amico, 2009; Valentine & McHugh, 2007). ... In the absence of reported attrition, reviewers can sometimes infer attrition rates from the research report [e.g., initial participants vs. degrees of freedom]" (p. 227). "Differential attrition among intervention groups or severe overall attrition is documented." "Is overall attrition less than 30%?" (p. 152).
6.9	"6.9. Differential attrition (between groups) is low (e.g., $\leq 10\%$) or is controlled for by adjusting for noncompleters (e.g., conducting intent-to-treat analysis)." "Given the lack of evidence linking differential attrition to ESs, we suggest reviewers use a relatively liberal criterion (e.g., $\leq 10\%$) for QI 6.9. QI 6.9 is met even when a study exhibits differential attrition exceeding reviewers' criterion when researchers demonstrate it had no impact (e.g., demonstrating pretest equivalence of groups after attrition) or is controlled for by using intent-to-treat analysis. If attrition rates for groups are not reported and cannot be inferred, QI 6.9 is not met, except when overall attrition is reported and is less than the review team's criterion for differential attrition" (p. 227). "Differential attrition among intervention groups or severe overall attrition is documented." "Researchers should document overall attrition of participants and ensure that attrition rates between the intervention and comparison groups were not substantially different" (p. 155).
7.0 Outcome Measures	7.0. Outcome measures/dependent variables. Outcome measures are applied appropriately to gauge the effect of the

	practice on study outcomes. Outcome measures demonstrate adequate psychometrics.
7.1	7.1. Outcomes are socially important (e.g., they constitute or are theoretically or empirically linked to improved quality of life, an important developmental/learning outcome, or both). A pre-intervention SV questionnaire is sufficient, or if a strong case is made for social significance in the introduction (e.g., there is a strong rationale and the intervention worked). The introduction can establish the social validity of the dependent variable (the social significance of the goal). If there is a measure of SV, score yes. If we can infer a meaningful change, score yes and note why. A meaningful change is defined as: "a functional relation between the introduction of a variable and change in the target behavior (e.g., reduction in challenging behavior)." "Although researchers should ideally make the case for the social validity of outcomes in their research report, QI 7.1 is also met when outcomes clearly constitute or are linked to improved quality of life or an important developmental or learning outcome" (p. 227). "Demonstration that the intervention produced an effect that met the defined, clinical need" (p. 172). "Demonstration that typical intervention agents (a) report procedures to be acceptable, (b) report the procedures to be feasible within available resources, (c) report the procedure to be effective, and (d) choose to continue use of the intervention procedures after formal support/expectation of use is removed" (p. 172)
7.2	"7.2. The study clearly defines and describes measurement of the dependent variables." "If a study measures multiple outcomes, reviewers should consider effects only for outcomes relevant to their review. For example, for a review of the effect of a practice on reading fluency, if a study measured effects of the practice on both reading fluency and reading comprehension, only the former effects should be considered. If a study examines

the effects of a practice on two or more measures of the outcome of focus (e.g., two measures of reading fluency), each dependent variable is treated as a separate case when determining the overall effect of the study. For example, in an ABAB study examining the impact of a practice on two different measures of reading fluency for three participants, meaningful, positive effects of the IV on the DV must occur for at least five of the six graphed outcomes ($\geq 75\%$) for the study to be categorized as having positive effects" (p. 8).

Study provides a logical explanation of the match between the dependent variable and data measurement system (e.g., whole interval, partial interval, duration, latency... and why). This QI is met if you could measure the same behavior using the same recording system.

"Providing a description of the measure and the process of administering/conducting the measure (or a reference to a publicly available source that describes the measure and its administration) in sufficient detail so that the reader could reasonably replicate the measurement procedure is necessary to meet QI 7.2. Merely mentioning the construct measured (e.g., spelling performance) without specifying the measure and its administration does not meet this QI" (p. 227).

Dependent variables are described with operational precision.

"Multiple measures are used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance."

"One of the foremost challenges in crafting a study is selection of measures that are well aligned with the substance and the intervention and yet are sufficiently broad and robust to (a) avoid criticism for "teaching to the test" through the specific intervention, and (b) demonstrate that generalizable skills have been successfully taught. An intervention may have adverse effects or additional benefits that a researcher should attempt to identify by using measures of generalized performance. Multiple measures should be used in a study, as no measure is perfect, and no measure can assess all or even most of the important aspects of performance that an intervention might affect" (p. 158).

7.3	7.3. The study reports the effects of the intervention on all measures of the outcome targeted by the review (p levels and effect sizes or data from which effect sizes can be calculated for group comparison studies; graphed data for single-subject studies), not just those for which a positive effect is found. "To avoid outcome reporting bias (Dwan et al., 2008), researchers must fully report findings for all the analyses and outcomes described in the research report for outcomes relevant to the review (QI 7.3). For example, for a review on the effects of a practice on math outcomes, reporting a p level and an ES for the effect on multiplication fluency but merely noting the analysis on solving word problems was insignificant (without a p value or an ES estimate) would not meet QI 7.3" (p. 227). "Were results presented in a clear, coherent fashion?"
7.4	7.4. Frequency and timing of outcome measures are appropriate. "QI 7.4 requires the outcome measures be conducted with appropriate timing and frequency. For group comparison studies, outcomes must be measured in reasonable proximity of the intervention. For example, using a pretest measure assessed 3 months before the intervention or a posttest measure assessed 3 months after the intervention poses a serious threat to validity and would not meet QI 7.4. Delays in posttesting are acceptable if done purposefully to investigate long-term effects of the practice. Because at least three data points are necessary to establish a trend, at least three data points must be measured for each phase in a single-subject study to be considered part of a possible demonstration of experimental effect (QI 6.5 requires at least three demonstrations of experimental effects at three different times). As alternating treatment designs do not involve comparisons between phases, to meet QI 7.4, they must involve four repetitions of the alternating sequence (Kratochwill et al., 2013)" (p. 227). "Outcomes for capturing the intervention's effect are measured at the appropriate times."

	"Were outcomes for capturing the intervention's effect measured beyond an immediate posttest?" (D) "Many times, the effects of an intervention are best detected immediately. ... If data collection is delayed, initial intervention effects may fade or subsequent instruction or other uncontrolled events may contaminate results. Sometimes, however, important intervention effects cannot be detected immediately. ... In cases such as this, researchers should time data collection activities to align with research questions. For some studies, it may be appropriate to collect data at only preand posttest. In many cases, however, researchers should consider collecting data at multiple points across the course of the study, including follow-up measures" (p. 159).
7.5	7.5. The study provides evidence of adequate internal reliability, interobserver reliability, test-retest reliability, or parallel-form reliability, as relevant (e.g., score reliability coefficient $\geq .80$, interobserver agreement $\geq 80\%$, kappa $\geq 60\%$). Reporting the average is okay, or if the range is below 80% but the average is 80% or higher, that's okay. As long as the low end of the range is not below 60% (or as otherwise agreed upon by the author research team conducting the systematic literature review). "Outcome measures should be both reliable and valid. QI 7.5 is met when any relevant type of reliability is assessed, reported, and shown to be adequate for outcomes relevant to the review. We recommend that reviewers use typical criteria for acceptable reliability (e.g., score reliability coefficient $\geq .80$, inter-observer agreement $\geq 80\%$, kappa $\geq 60\%$). Study authors must report a specific coefficient or agreement level. Noting that "disagreements were reconciled" or "reliability was found to be acceptable" does not address QI 7.5. QI 7.5 does not apply to extremely low-inference measures. For example, graduation status as reported by the school does not require evidence of reliability when used to measure graduation. ... QIs 7.5 and 7.6 can also be addressed by citing previously published reports. If evidence of reliability or validity is imported from another study, the sample and scores must be similar enough to make generalization to the current study sensible.

	Adequacy of score and sample comparability can be indicated in the research report or determined by reviewers, which assumes the cited source is publicly available. If reliability or validity is established solely by citing another source that is not publicly available, the QI is not met" (p. 228). "Evidence of reliability and validity for the outcome measures is provided." "Adequate interscorer agreement is documented." "Data collectors and/or scorers are blind to study conditions and equally (un)familiar to examinees across study condition." (D) "Did the study provide not only internal consistency reliability but also test–retest reliability and interrater reliability (when appropriate) for outcome measures?" (p. 152). "The bottom line for reliability are coefficient alpha reliabilities of at least .6" (p. 159). "Ideally, reliability in data collection and scoring should be above .90" (p. 160) "Researchers should design and carry out data collection activities in a way to minimize these threats to the study's internal validity" (p. 160)
7.6	The study provides adequate evidence of validity, such as content, construct, criterion (concurrent or predictive), or social validity. "Validity can be assessed in multiple ways, and commonly accepted, uniform standards do not exist for validity. For example, standards for concurrent validity coefficients tend to be lower in the area of writing than for other academic outcomes (e.g., McMaster, Du, & Pétursdóttir, 2009), content validity is typically determined by judgment, and construct validity is defined and examined in multiple ways. QI 7.6 is met by (a) study authors reporting adequate validity coefficients (as determined a priori by the review team) or (b) the outcome adequately representing the content being measured (i.e., content validity; as justified by study authors or as determined by reviewers). Adequate

	content validity can be assumed when a discrete behavior, whether a microbehavior, such as tying shoes, or a macrobehavior, such as graduating, is measured" (p. 228). "Was evidence of the criterion-related validity and construct validity of the measures provided?" (D)
8.0 Data Analysis	8.0. Data Analysis. Data analysis is conducted appropriately. The study reports information on effect size.
8.1	8.1. Data analysis techniques are appropriate for comparing change in performance of two or more groups (e.g., t tests, ANOVAs/MANOVAs, ANCOVAs/MANCOVAs, hierarchical linear modeling, structural equation modeling). If atypical procedures are used, the study provides a rationale justifying the data analysis techniques. "QI 8.1 is met when (a) researchers use statistical analyses generally recognized as appropriate for comparing change in the performance of two or more groups (e.g., t tests, ANOVAs/MANCOVAs, hierarchical linear modeling, structural equation modeling) or (b) atypical procedures are used but justified by researchers. Justification for atypical analytic procedures can be provided through publicly available references. Using analyses with which special education researchers with advanced research training are unfamiliar and that are not explained does not meet this QI" (p. 228). "The data analysis techniques chosen are appropriate and linked in an integral fashion to key research questions and hypotheses." "The researcher should clearly link the unit of analysis chosen to the key statistical analyses." "Statistical analyses are accompanied with presentation of effect sizes. Merely cataloguing every possible analysis that can be done is unacceptable. A brief rationale for major analyses and for selected secondary analyses is critical" (p. 161).
8.3	8.3. The study reports one or more appropriate effect size statistic (e.g., Cohen's d, Hedge's G, Glass's Δ , η^2) for all outcomes relevant to the review being conducted, even if

	the outcome is not statistically significant, or provides data from which appropriate effect sizes can be calculated. "To meet QI 8.3, researchers must report an appropriate ES statistic (e.g., Cohen's d, Hedge's G, Glass's Δ, η^2) or provide data from which a common ES statistic can be calculated (e.g., group means and standard deviations, t values and sample sizes, F values and degrees of freedom) for each analysis examining the targeted practice's effect in group comparison studies. QI 8.3 is also met when study authors report atypical ESs but describe and justify their use or cite a publicly available source justifying their use" (p. 228). "Did the research report include not only inferential statistics but also effect size calculations?" "Effect sizes in the .40 or larger range are often considered minimum levels for educational or clinical significance" (p. 161).
--	---

Content retrieved and adapted from:

Cook, B. G., Tankersley, M., & Landrum, T. J. (2015). The CEC 2014 standards for evidence-based practices in special education. *Council for Exceptional Children*. These standards are referenced in the context of evaluating the methodological soundness of studies and determining evidence-based practices in special education.

Council for Exceptional Children. (2014). *Exceptional Children*, 80(4), 504–511.

<https://doi.org/10.1177/0014402914531388>

Horner, R. H., Carr, E. G., Halle, J., McGee, 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. This article discusses the defining features of single-subject research and its role in establishing evidence-based practices in special education.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005).

Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149–164. This work is widely cited for its discussion of research quality indicators and the challenges of using randomized controlled trials in special education

Kennedy, M. J. (2005). Single-case designs for educational research. *Council for Exceptional Children*. Kennedy's work is referenced in the context of the importance and application of single-case research in special education.