PsychCorp

Volume 10, Issue 2 September 2015

EBP briefs

A scholarly forum for guiding evidence-based practices in speech-language pathology

Telepractice vs. On-Site Treatment: Are Outcomes Equivalent for School-Age Children?

Johanna M. Rudolph
University of Texas at Dallas, Dallas, TX

Stephen Rudolph
Independent Scholar, Dallas, TX

ALWAYS LEARNING

PEARSON

EBP Briefs

Cite this document as:
Rudolph, J. M. & Rudolph, S. (2015). Telepractice vs. on-site treatment: are outcomes equivalent for school-age children? *EBP Briefs, 10*(2), 1–15. Bloomington, MN: NCS Pearson, Inc.

PsychCorp

PsychCorp is an imprint of Pearson Clinical Assessment.

Pearson Executive Office 5601 Green Valley Drive Bloomington, MN 55437 800.627.7271 www.PearsonClinical.com

Produced in the United States of America.

1 2 3 4 5 6 7 8 9 10 11 12 A B C D E

PEARSON

Structured Abstract

Clinical Question: Do school-age children receiving treatment for speech, language, and/or communication disorders show equivalent benefit from telepractice-based intervention as from on-site intervention as shown by comparable improvement in speech, language, and communication skills across the two treatment platforms?

Method: Systematic Review and Meta-Analysis

Study Sources: PubMed, SCOPUS, CINAHL, ERIC, MEDLINE, PsycINFO, and Mendeley

Search Terms: telepractice OR telehealth OR telemedicine OR telerehabilitation OR telecare OR telespeech AND speech OR language OR communication AND interven* OR treat* OR therap* AND child*

Number of Included Studies: 6

Primary Results:

All studies reported equivalent or greater improvement in the telepractice group; however, the confidence intervals associated with the effect sizes were large.

Effect size was negatively correlated with study quality with studies of poorer quality reporting larger effects (in favor of telepractice) and studies of higher quality reporting smaller or negative effects (in favor of on-site).

Meta-analysis of the effect sizes from the three studies of highest quality yielded a wide confidence interval (-0.56 to 0.20) suggesting that the true effect size could range from a moderate effect in favor of on-site to a small effect in favor of telepractice.

Conclusions: Although telepractice appears to be a promising platform for delivering speech and language services to school-age children, there is not yet sufficient evidence to confirm that outcomes resulting from treatment provided via telepractice are equivalent to those resulting from on-site treatment. Future studies exploring this question should focus on obtaining adequate sample sizes to meet power requirements, employing suitable controls, and implementing other quality metrics to ensure the validity and reliability of results.

Telepractice vs. On-Site Treatment: Are Outcomes Equivalent for School-Age Children?

Johanna M. Rudolph
Stephen Rudolph

Clinical Scenario

Denise, a certified speech-language pathologist (SLP), has worked full-time as an SLP in her local independent school district for the past five years. Toward the end of the school year, her director notified her that the SLP at a rural school within her district made plans to move out of state leaving an empty position at the school and a caseload of kids requiring services under the Individuals with Disabilities Education Act (IDEA, 2004). The school will attempt to fill the position in the few months before the start of the next school year, but it is likely that another SLP in the district will have to fill in for a period of time. The director explained that, as one of the most proximate SLPs, Denise should expect to take on extra students. Denise calculated how many additional hours she would need to dedicate to her school position in the upcoming year and found that travel time alone would increase about 2.5 hours every week (travel to and from the rural school at least twice per week). Denise began exploring ways she could increase efficiency while still providing high-quality services to all of the students on her caseload. A brief search of the American Speech-Language-Hearing Association (ASHA) Schools Professional Issues website (http://www.asha.org/SLP/schools/Resources-for-school-based-SLPs) revealed that telepractice might be the perfect solution because it would remove the burden of commuting to the rural school. She consulted with the SLP director to determine whether telepractice was a viable option. The director requested, as a first step, that she review the available evidence to determine how therapy provided via telepractice compares to on-site therapy. Denise agreed readily to the task.

Background Information

Telepractice is defined as "the application of telecommunications technology to the delivery of speech-language pathology and audiology professional services at a distance by linking clinician to client/patient or clinician to clinician for assessment, intervention, and/or consultation" (ASHA, n.d.). Although telepractice is the term preferred and used by ASHA, many other terms have been applied to the provision of services through videoconference or related technology. These include telehealth, telemedicine, telerehabilitation, and telecare among others and these services have been provided in a wide variety of settings (see McLean et al., 2013 for a review).

Denise searched first for review articles related to her topic of interest. She found reviews focused on the use of telepractice for children with autism spectrum disorders (ASD) (Boisvert, Lang, Andrianopoulos, & Boscardin, 2010) and children with fluency disorders (Lowe, O'Brian, & Onslow, 2013). She also found a review focusing on pediatric speech and language assessment via telepractice (Taylor, Armfield, Dodrill, & Smith, 2014) and several papers reviewing telepractice generally in the field of speech-language pathology (Hill & Theodoros, 2002; Mashima & Doarn, 2008; Reynolds, Vick, & Haak, 2009; Theodoros, 2011). These reviews indicated that telepractice is a promising approach for delivering SLP services to individuals with neurogenic disorders, voice disorders, stuttering, dysphasia, ASD, and other developmental disorders. However, Denise found no reviews that directly addressed the topic of telepractice for the provision of these services in the schools. This was surprising given that schools are the most common setting in which telepractice services are delivered (ASHA, n.d.). Denise realized that she needed to complete her own review to make an evidence-based decision about the implementation of telepractice in her school district. Systematic review and meta-analysis procedures were used to summarize the literature pertaining to the research question.

Clinical Question

To guide her literature search, Denise formulated her clinical question using the PICO model (Oxford Centre for Evidence Based Medicine, University of Oxford, 2014) where P stands for the patient, population, or problem under consideration; I stands for the intervention or exposure; C stands for the comparison intervention; and O stands for the outcome. Her PICO terms included:

P – school-age children who receive treatment for speech, language, and/or communication disorders

I – telepractice-based intervention

C – on-site intervention

O – comparable improvement in speech, language, and/or communication skills.

The question derived from these terms was: Do school-age children receiving treatment for speech, language, and/or communication disorders show equivalent benefit from telepractice-based intervention as from on-site intervention as shown by comparable improvement in speech, language, and communication skills across the two treatment platforms? Denise immediately recognized that her question differed from most other PICO questions in that she wanted to establish whether the experimental intervention (telepractice) was equivalent to instead of better than the standard (on-site).

Search for the Evidence

Search Strategy. After establishing her clinical question, Denise chose the following search string to identify relevant literature: telepractice OR telehealth OR telemedicine OR telerehabilitation OR telecare OR telespeech AND speech OR language OR communication AND interven* OR treat* OR therap* AND child*. She completed the search in May 2015 using databases that were available to her due to a university affiliation: PubMed, SCOPUS, CINAHL, ERIC, MEDLINE, and PsycINFO. She also used the search function in her externally-linked reference organization software (Mendeley). Denise excluded books and conference proceedings from her citation tally. The search yielded 703 citations, 398 of which were unique.

Inclusion Criteria. Denise organized her citations using Microsoft® Excel®. She evaluated them in three phases: by title, by abstract, and by full text. She used the following inclusionary criteria to guide her evaluation: the study must (a) be available in English; (b) be focused on telepractice; (c) be focused on treatment of speech, language, or communication disorders; (d) include preschool or school-age children [i.e., children treated under an Individualized Education Plan (IEP)]; and (e) employ an intervention that would/could be implemented in a school setting. When assessing full texts, two additional criteria were considered: the study must (f) include speech, language, or communication outcome data; and (g) include an on-site comparison group. When review articles that met the first five criteria were identified, their reference lists were searched for further relevant studies. This yielded an additional 37 citations, which were evaluated using the same procedures. At each phase, citations were only eliminated when an explicit reason for exclusion could be identified. Vague references were retained for further evaluation. Figure 1 displays Denise's decision process at each stage. Six articles qualified for the review.

To confirm the results of this process, Denise asked a colleague to evaluate approximately 30% of the citations (n = 132) using the specified inclusionary criteria. At the title, abstract, and full text phases, agreement was 88%, 81%, and 100%, respectively. Discrepancies were discussed and she and her colleague resolved to the original decision in all but four instances. Denise interpreted this outcome as confirmation of the accuracy of her assessment.

Evaluating the Evidence

Extraction of Study Characteristics. As a first step in the process of evidence evaluation, Denise extracted the following characteristics from each study: (a) sample size; (b) research design; (c) participants' age and diagnoses; (d) inclusion criteria; (e) service delivery characteristics (length and frequency of sessions, duration of intervention, service delivery method, location); (f) treatment characteristics (approach, targets); and (g) details of telepractice implementation (hardware, software, connection, privacy). Denise also calculated effect sizes for the speech, language, and communication outcomes reported. The effect of interest, given the PICO question, was the difference in progress (or change) from pretest to posttest for children receiving treatment via telepractice versus those receiving treatment on-site. As such, effect sizes could only be calculated when data reflecting pretest–posttest change were available. Effect sizes were derived using the calculator from The Campbell Collaboration website (http://www.campbellcollaboration.org/resources/effect_size_input.php), which provides values and confidence intervals for Cohen's d (Cohen, 1988). When more than one outcome measure was reported within a study, Denise calculated a combined effect using Comprehensive Meta-Analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2014). Thus, each study was represented by a single effect size to increase the ease of comparison and to avoid the problem of assigning more weight to studies that reported more outcomes (Borenstein,

Hedges, Higgins, & Rothstein, 2008). Study characteristics and effect sizes are summarized in Appendix A.

Findings. Denise found that there were a total of 187 participants, ages 3 to 15 years, across the six studies. Among these, 122 were assigned to telepractice (or telepractice first) groups, 49 were assigned to on-site (or on-site first) groups, and 16 were assigned to a placebo group. The placebo group received online training as opposed to direct training from a qualified specialist. Because this review was focused on direct training, the placebo group was not considered in the analysis phase. Note that Gabel, Grogan-Johnson, Alvares, Bechstein, & Taylor (2013) used a database sample of 5,332 children as their comparison group (not included in participant count). Participant diagnoses included ASD, speech sound disorder, language impairment, learning disability, and fluency disorder. Only one study mentioned that participants had to meet ASHA criteria for telepractice candidacy in order to qualify (Boisvert, 2012). The length and frequency of sessions varied, but generally fell in the range of 20 to 30 minutes once or twice per week. The one study that did not fall in this range focused on training special-education teachers to increase the social and communicative function of individuals with ASD (Ruble, McGrew, Toland, Dalrymple, & Jung, 2013). Despite the length of the sessions in this study (90 min), they only occurred twice per semester, which is a feasible service delivery model for a school SLP due to the infrequency of the sessions. Most telepractice sessions were administered individually and in a pull-out context. In all but one study (Grogan-Johnson et al., 2013), participants were located at their assigned schools during treatment sessions. Treatment approaches and treatment targets varied substantially, but speech sounds were the most common target across the studies. In order to provide sufficient audio and video quality, the studies were all conducted with computer hardware and network connections that met or exceeded the specifications recommended by their chosen videoconferencing software platform. The chosen platforms were a mix of freemium (one used Skype's™ free tier) and commercial software (three used Polycom® PVX™ and one used Adobe® Connect Pro). The studies using commercial software did not specify whether they were chosen because of existing licensing arrangements between their institution and the chosen vendor or for reasons directly related to the study. In an attempt to comply with the Health Insurance Portability and Accountability Act (HIPAA, 1996) regulations, five of the six studies used computer-based videoconferencing software that provided encrypted communication with a media server. In all but one case (Ruble et al., 2013) a commercial third party hosted the media server, which has additional compliance requirements under HIPAA. The sixth study omitted any mention of this consideration. Effect sizes varied substantially from study to study.

Extraction of Study Quality Features. In order to assess study quality, Denise referenced a checklist downloaded from the Oxford Centre for Evidence-Based Medicine (CEBM) website (Oxford Centre for Evidence-Based Medicine, University of Oxford, 2005). Based on this checklist, she decided to evaluate the studies using the following constructs: (a) study design, (b) internal validity, (c) external validity, and (d) reliability. Due to the unique focus of her PICO question, Denise added a fifth consideration: whether the studies had adequate (e) power to detect a small effect. If not, findings of equivalence between telepractice and on-site treatment could be attributable to Type II error (rejection of an effect when an effect actually exists) rather than true equivalence.

To assess study design, Denise used the CEBM levels of evidence (OCEBM Levels of Evidence Working Group, 2011). To assess internal validity, Denise considered whether the experimental and control groups were similar in diagnosis, pre-intervention characteristics, type of intervention received, number of sessions received, service delivery method (pull-out, group, individual, etc.), and goals targeted. Two additional internal validity considerations were objectivity of the outcome measures and blinding of assessors. To assess external validity, she determined whether the participants were similar to those students one might typically encounter in a school SLP's caseload and whether the service delivery model (length, frequency, and method) was feasible in a school setting. To assess reliability, Denise examined interrater reliability scores and treatment integrity data. To assess power, Denise first established whether a power analysis had been performed by the authors and then examined the sample size to determine whether it was large enough to meet power requirements.

Upon initial review of the studies, Denise noted substantial variability in quality. She was concerned that these differences in methodological rigor might introduce significant interstudy heterogeneity, which would preclude aggregation of effect sizes through meta-analysis. Denise resolved to examine the association between study quality and effect size using random effects meta-regression. In

order to implement this procedure, she needed to assign numeric quality ratings to the studies. She looked for an appraisal tool that evaluated all of the specified features and provided a summary score, but she was unable to find one that met the unique requirements of her review corpus, so Denise developed a numerical coding protocol based on the principle that the five major constructs (i.e., power, design, internal validity, external validity, reliability) should be equally weighted. Each construct was assigned a maximum value of 20 for a possible maximum total score of 100. This protocol is outlined in Appendix B. Denise was aware of the pitfalls of using a scoring system including the somewhat arbitrary assignment of point values for different features and the different types of bias that must be represented by a single value. She had been attentive to these considerations when developing the protocol, and she anticipated that a significant association between effect size and study quality in the meta-regression analysis would reinforce the construct validity of the appraisal tool. If she found no such association, study quality could still be summarized narratively based on the scoring system.

Findings. The results of the coding procedure are reported in Table 1. Quality scores ranged from 37 to 65. The construct that was most often unaddressed (or under-addressed) was power. Only one study completed a power analysis (Ruble et al., 2013), but the authors reported that they were unable to recruit enough participants to meet power requirements. A second study (Gabel et al., 2013) did not complete a power analysis, but received points for power because the sample size was most likely sufficient to detect a small effect. Other constructs that were rarely addressed included assessor blinding, treatment integrity, and interrater reliability. The absence of these quality control features decreased the internal validity of the associated studies and the reliability of the results they reported. However, the external validity for all of the studies was high.

The Evidence-Based Decision

In order to answer her PICO question, Denise proceeded with her planned meta-analyses. First, she performed meta-regression using quality score as the moderator variable and effect size as the dependent variable. This analysis indicated that when study quality was not taken into consideration, significant levels of heterogeneity existed among the effect sizes (Q = 14.32, p = 0.01). However, when study quality was taken into consideration, heterogeneity was eliminated (Q = 7.19, p = 0.13). This was due to the fact that quality score was significantly and inversely related to effect size (coefficient = -0.06, SE = 0.03, Z = -2.22, p = 0.03), that is, the higher the quality of the study, the smaller the effect size (see Figure 2). This finding supported the validity of her appraisal tool. Given this result, Denise decided it would be most informative to base her decision on the three studies that received quality scores over 50 (indicating that more than 50% of the quality metrics were in place). She found there was no significant heterogeneity across these studies (Q = 0.63, p = 0.73) and entered the effect sizes into meta-analysis. The aggregate effect size was small and nonsignificant (d = -0.18, 95% CI = -0.56–0.20). Based on the guidelines for interpretation of Cohen's d (0.20 = small effect, 0.50 = moderate effect, 0.80 = large effect), an effect size of -0.18 suggests there is little to no difference in the performance of children treated via telepractice versus those treated on-site. However, when Denise examined the confidence interval, she found that the true effect size could be any value between -0.58 (indicating a moderate effect in favor of on-site treatment) and 0.20 (indicating a small effect in favor of telepractice treatment). Denise surmised that this uncertainty was due to the fact that there were so few studies of reasonable methodological quality available to answer her question. Denise concluded that she could not confidently say that speech and language treatment provided to school-age children via telepractice is equivalent to services provided on-site.

When Denise presented the results of her review to the director, he agreed that the evidence was not sufficient to warrant implementation of a telepractice program in their district, particularly since on-site treatment was a viable option. A district meeting was subsequently scheduled to discuss the possibility of dividing the rural school caseload among multiple clinicians to ensure that the impact on any single SLP was minimized. Denise was grateful for the assistance and hopeful that future high-quality research focusing on the use of telepractice in schools would shed light on her still unanswered question.

Authors' Note

Johanna Rudolph is a postdoctoral fellow at the University of Texas at Dallas Callier Center for Communication Disorders. Her primary research interests include early identification and intervention for children with speech and language disorders.

Johanna M. Rudolph
Callier Center for Communication Disorders
University of Texas at Dallas
Dallas, TX 75235
johanna.rudolph@utdallas.edu

Stephen Rudolph is an independent scholar formerly affiliated with the Purdue University Rendering and Perceptualization Lab. Currently, his interests include helping researchers improve the efficiency and effectiveness of their work through the use of software tools and programming skills.

References

Papers marked with * were included in the review.

American Speech-Language-Hearing Association. (n.d.). *Telepractice.* (Practice Portal). Retrieved from www.asha.org/Practice-Portal/Professional-Issues/Telepractice

*Boisvert, M. K. (2012). *An investigation of the efficacy of speech and language interventions with students with ASD using telepractice.* (Doctoral dissertation). Retrieved from http://scholarworks.umass.edu/open_access_dissertations/536

Boisvert, M., Lang, R., Andrianopoulos, M., & Boscardin, M. L. (2010). Telepractice in the assessment and treatment of individuals with autism spectrum disorders: A systematic review. *Developmental Neurorehabilitation, 13*(6), 423–432. doi:10.3109/17518423.2010.499889

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2008). *Introduction to meta-analysis.* Hoboken, NJ: John Wiley & Sons, Inc.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive meta-analysis (Version 3)* [Computer software]. Englewood, NJ: Biostat.

Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

*Gabel, R., Grogan-Johnson, S., Alvares, R., Bechstein, L., & Taylor, J. (2013). A field study of telepractice for school intervention using the ASHA NOMS K-12 database. *Communication Disorders Quarterly, 35*(1), 44–53. doi:10.1177/1525740113503035

*Grogan-Johnson, S., Alvares, R., Rowan, L., & Creaghead, N. (2010). A pilot study comparing the effectiveness of speech language therapy provided by telemedicine with conventional on-site therapy. *Journal of Telemedicine and Telecare, 16*(3), 134–139. doi:10.1258/jtt.2009.090608

*Grogan-Johnson, S., Gabel, R. M., Taylor, J., Rowan, L. E., Alvares, R., & Schenker, J. (2011). A pilot investigation of speech sound disorder intervention delivered by telehealth to school-age children. *International Journal of Telerehabilitation, 3*(1), 31–42. doi: 10.5195/ijt.2011.6064

*Grogan-Johnson, S., Schmidt, A. M., Schenker, J., Alvares, R., Rowan, L. E., & Taylor, J. (2013). A comparison of speech sound intervention delivered by telepractice and side-by-side service delivery models. *Communication Disorders Quarterly, 34*(4), 210–220. doi:10.1177/1525740113484965

Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1938 (1996).

Hill, A., & Theodoros, D. (2002). Research into telehealth applications in speech-language pathology. *Journal of Telemedicine and Telecare, 8*(4), 187–196. doi:10.1258/135763302320272158

Individuals with Disabilities Education Act of 2004, Pub. L. No. 108-446, § 101, 118 Stat. 2647 (2004).

Lowe, R., O'Brian, S., & Onslow, M. (2013). Review of telehealth stuttering management. *Folia Phoniatrica et Logopaedica, 65*(5), 223–238. doi:10.1159/000357708

Mashima, P. A., & Doarn, C. R. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and E-Health, 14*(10), 1101–1117. doi:10.1089/tmj.2008.0080

McLean, S., Sheikh, A., Cresswell, K., Nurmatov, U., Mukherjee, M., Hemmi, A., & Pagliari, C. (2013). The impact of telehealthcare on the quality and safety of care: A systematic overview. *PLoS ONE, 8*(8), e71238. doi:10.1371/journal.pone.0071238

OCEBM Levels of Evidence Working Group. (2011). The Oxford 2011 levels of evidence. *Oxford Centre for Evidence-Based Medicine.* Retrieved from http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf

Oxford Centre for Evidence-Based Medicine, University of Oxford. (2005). *Critical appraisal worksheets: Therapy/RCT articles.* Retrieved from http://www.cebm.net/critical-appraisal

Oxford Centre for Evidence-Based Medicine, University of Oxford. (2014). *Asking focused questions.* Retrieved from http://www.cebm.net/asking-focused-questions

Reynolds, A. L., Vick, J. L., & Haak, N. J. (2009). Telehealth applications in speech-language pathology: A modified narrative review. *Journal of Telemedicine and Telecare, 15*(6), 310–316. doi:10.1258/jtt.2009.081215

*Ruble, L. A., McGrew, J. H., Toland, M. D., Dalrymple, N. J., & Jung, L. A. (2013). A randomized controlled trial of COMPASS web-based and face-to-face teacher coaching in autism. *Journal of Consulting and Clinical Psychology, 81*(3), 566–572. doi:10.1037/a0032003

Taylor, O. D., Armfield, N. R., Dodrill, P., & Smith, A. C. (2014). A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *Journal of Telemedicine and Telecare, 20*(7), 405–412. doi:10.1177/1357633X14552388

Theodoros, D. (2011). Telepractice in speech-language pathology: The evidence, the challenges, and the future. *Sig 18 Perspectives on Telepractice, 1*(1), 10–21. doi:10.1044/tele1.1.10

Table 1. Quality Score Coding and Results for Included Studies

Quality metric		Studies					
		Boisvert (2012)	Gabel et al. (2013)	Grogan-Johnson et al. (2010)	Grogan-Johnson et al. (2011)	Grogan-Johnson et al. (2013)	Ruble et al. (2013)
Power		0	20	0	0	0	0
Design		8	12	16	12	16	16
Internal validity	Diagnosis	2	0	0	2	2	2
	Matching	0	1	3	3	3	2
	Treatment	2	0	0	2	2	2
	Sessions	0	0	0	2	2	2
	Method	2	0	0	0	2	2
	Targets	0	1	0	1	1	0
	Outcomes	3	0	3	3	3	0
	Blinding	0	0	0	0	2	4
	SUBTOTAL	9	2	6	13	17	14
External validity	Population	10	10	10	10	10	10
	Model	5	10	5	5	5	5
	SUBTOTAL	15	20	15	15	15	15
Reliability	IRR	10	0	0	5	0	10
	TI	0	0	0	0	10	10
	SUBTOTAL	10	0	0	5	10	20
Total Quality Score		**42**	**54**	**37**	**45**	**58**	**65**

Note. Maximum possible score for each major construct (i.e., power, design, internal validity, external validity, reliability) is 20. Studies received 0 points for a given feature if information was not provided and could not be determined. IRR = Interrater Reliability, TI = Treatment Integrity.

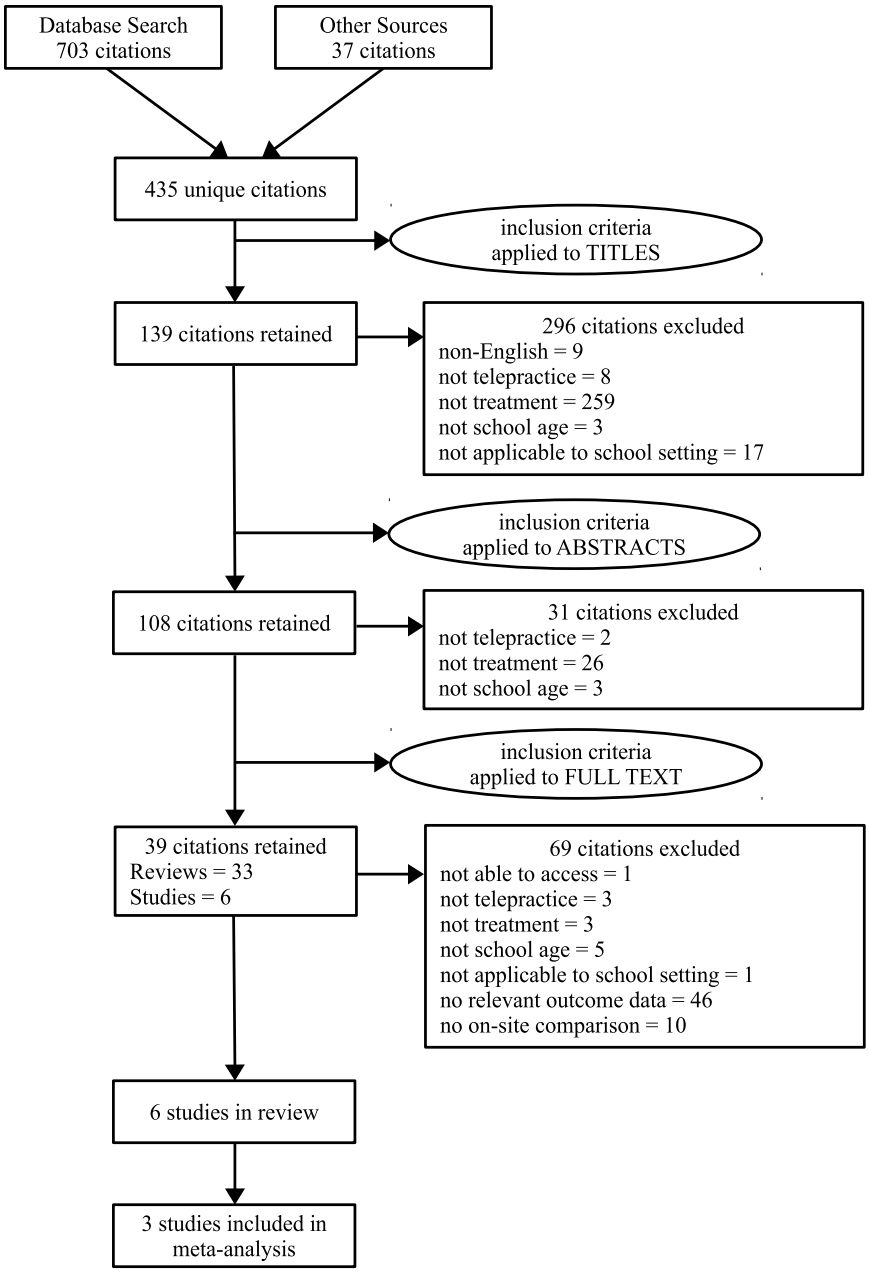

Figure 1. Results of the study search and three phase evaluation process.

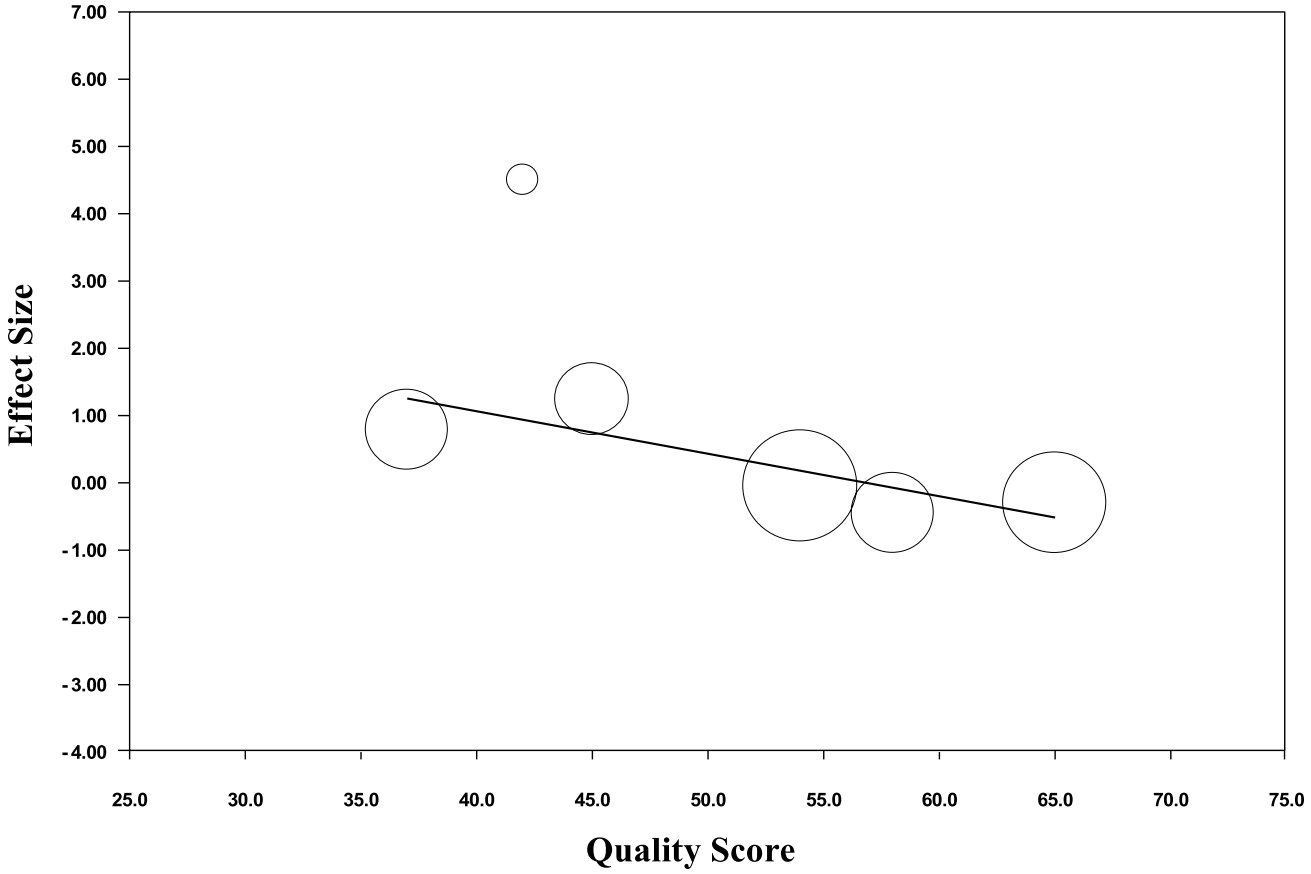

Figure 2. Results of the meta-regression analysis examining the association between study quality (quality score) and difference in improvement between telepractice and on-site groups (effect size). Quality score accounted for 56% of the variance in effect size.

Appendix A. Characteristics and Effect Sizes for Studies Included in Systematic Review

Study	**Boisvert, 2012**		
Sample Size	*N* = 6 (5 male, 1 female); 3 telepractice first, 3 on-site first		
Design	Single subject (Level 4); crossover trial		
Participants	5–12 years; diagnosed with ASD *Inclusion criteria:* must participate in mainstream class 80% of each day, communication goals on IEP, English as primary language, good health status, at least 50 words (10 oral), must meet ASHA criteria for telepractice candidacy, no other primary diagnosis, no uncorrected sensory deficits, no recent history of property destruction, no recent history of injury to self or others		
Service Delivery	30 min; 1–2x per week; 12 weeks (6 weeks per phase) Individual pull-out; school setting		
Treatment	*Intervention:* structured trials and naturalistic trials *Target (differed by student):* social engagement, answering questions, following directions, transition words, grammatical structures, speech sounds, vocabulary/concepts		
Telepractice Specifications	*Hardware* • clinician: not specified • student: computer with speakers, external webcam with embedded microphone *Software:* Skype™, Adobe® Connect Now, PresenceLearning *Connection:* not specified *Privacy:* encrypted connection to commercial servers		
Effect Size	**Outcome**	**Cohen's *d***	**95% Confidence Interval**
	Based on probe performance (unit = probe response)		
	Phase I probes (Δ from baseline)	4.50[b]	1.19–7.81
	Phase II probes (Δ from baseline)	1.99	–0.19–4.16
Study	**Gabel et al., 2013**		
Sample Size	*N* = 71 (45 male, 26 female); 71 telepractice, 5,332 on-site (database sample)		
Design	Nonrandomized controlled trial (Level 3)		
Participants	5–15 years; diagnosed with speech impairment, language impairment, learning disorder *Inclusion criteria:* no autism, no cognitive impairment, no cerebral palsy, no cleft lip/palate, no neurological impairment, no significant hearing loss, no significant visual impairment		
Service Delivery	~ 10 hours total across the academic year Individual pull-out, group pull-out, self-contained, collaborative consultative; school setting		
Treatment	*Intervention:* not specified *Target (differed by student):* intelligibility, fluency, pragmatics, speech sounds, language comprehension, language production, reading comprehension, voice		
Telepractice Specifications	*Hardware* • clinician: desktop, webcam with built-in microphone, headset • student: desktop, webcam with built-in microphone, headset • facilitator: headset *Software:* Polycom® PVX™ *Connection:* broadband internet *Privacy:* encrypted connection to commercial servers		

Appendix A. Characteristics and Effect Sizes for Studies Included in Systematic Review (continued)

Effect Size	**Outcome**	**Cohen's *d***	**95% Confidence Interval**
	Based on Functional Communication Measure improvement (unit = participant)		
	Speech sounds	0.23[a]	–0.25–0.71
	Language production	–0.35[a]	–0.81–0.11
	Language comprehension	–0.17[a]	–0.70–0.37
	Intelligibility	0.10[a]	–0.49–0.70
	Combined effect	–0.05[b,c]	–0.56–0.46
Study	**Grogan-Johnson et al., 2010**		
Sample Size	*N* = 34 (25 male, 13 female); 17 telepractice first, 17 on-site first		
Design	Randomized controlled trial (Level 2); crossover trial		
Participants	4–12 years; diagnosed with articulation disorder, language disorder, fluency disorder, learning disorder *Inclusion criteria:* current IEP, no autism or PDD, no severe cognitive deficit, no severe emotional disturbance		
Service Delivery	8 months (4 months per phase) Individual pull-out (telepractice), group pull-out (on-site); school setting		
Treatment	*Intervention:* not specified *Target (differed by student):* intelligibility, speech sounds, language production		
Telepractice Specifications	*Hardware* • clinician: computer, headphones, document camera • student: computer, headphones • facilitator: headphones *Software:* not specified *Connection:* broadband internet *Privacy:* not specified		
Effect Size	**Outcome**	**Cohen's *d***	**95% Confidence Interval**
	Based on GFTA-2 improvement (unit = participant)		
	Phase I GFTA-2 increase	0.20[a]	–0.72–1.13
	Phase II GFTA-2 increase	–0.44	–1.42–0.54
	Based on IEP goal progress (unit = goal)		
	Phase I adequate progress/mastered	1.37[a]	0.19–2.55
	Phase II adequate progress/mastered	0.18	–0.42–0.77
	Combined effect	0.79[b]	–0.27–1.84
Study	**Grogan-Johnson et al., 2011**		
Sample Size	*N* = 13 (11 male, 2 female); 7 telepractice, 6 on-site		
Design	Nonrandomized controlled trial (Level 3)		
Participants	6–11 years; diagnosed with speech sound disorder *Inclusion criteria:* communication impairment, IEP with goals for speech sound disorder, no autism or PDD, no cognitive deficits, no severe emotional disturbance, no visual impairment, no hearing impairment, no ESL students		
Service Delivery	20 min; ~2x per week; 6 months Individual pull-out; school setting		

Appendix A. Characteristics and Effect Sizes for Studies Included in Systematic Review (continued)

Treatment	*Intervention:* traditional articulation approach *Target:* speech sounds		
Telepractice Specifications	*Hardware* • clinician: desktop, webcam with built-in microphone, headset • student: desktop, webcam with built-in microphone, headset • facilitator: headset *Software:* not specified *Connection:* broadband internet *Privacy:* encrypted connection to commercial servers		
Effect Size	**Outcome**	**Cohen's *d***	**95% Confidence Interval**
	Based GFTA-2 performance (unit = probe response)		
	GFTA-2 (Δ from baseline)	2.09[a]	0.74–3.45
	Based on sound probe performance (unit = probe response)		
	Sound probe (Δ from baseline)	0.80[a]	–0.64–2.23
	Based on goal progress (unit = goal)		
	Mastered	0.83[a]	–0.01–1.67
	Combined effect	1.24[b]	0.00–2.48
Study	**Grogan-Johnson et al., 2013**		
Sample Size	*N* = 14 (9 male, 5 female); 7 telepractice, 7 on-site		
Design	Randomized controlled trial (Level 2)		
Participants	6–10 years; diagnosed with speech sound disorder *Inclusion criteria:* no significant hearing loss, no significant visual impairment, no cerebral palsy, no cognitive impairment, no cleft lip/palate, no neurological impairment, English as primary language		
Service Delivery	30 min; 2x per week; 5 weeks Individual; university clinic setting		
Treatment	*Intervention:* traditional articulation approach *Target:* speech sounds		
Telepractice Specifications	*Hardware* • clinician: desktop, webcam with built-in microphone, headset • student: laptop, webcam with built-in microphone, headset *Software:* Polycom® PVX™ *Connection:* broadband internet *Privacy:* encrypted connection to commercial servers		
Effect Size	**Outcome**	**Cohen's *d***	**95% Confidence Interval**
	Based on GFTA-2 subtest performance (unit = probe response)		
	GFTA-2 (Δ from baseline)	–0.45[b,c]	–1.51–0.61
	Based on listener judgment of sound accuracy (unit = speech sound)		
	Listener judgment (Δ from baseline)	0.45	–0.65–1.56

Appendix A. Characteristics and Effect Sizes for Studies Included in Systematic Review (continued)

Study	**Ruble et al., 2013**		
Sample Size	*N* = 49 special education teachers assigned to 49 students (42 male, 7 female); 17 telepractice, 16 on-site, 16 placebo		
Design	Randomized controlled trial (Level 2)		
Participants	3–9 years; diagnosed with ASD *Inclusion criteria:* diagnosis confirmed with ADOS, special services designated in IEP		
Service Delivery	3-hour introductory session (1x); 90 min coaching session (4x) every 5 weeks; academic year Individual coaching; school setting		
Treatment	*Intervention:* COMPASS *Target (differed by student):* communication, social skills, independence		
Telepractice Specifications	*Hardware* • clinician: not specified • student/teacher: laptop, webcam, headphones, video camera *Software:* Adobe® Connect Pro *Connection:* not specified *Privacy:* encrypted connection to a university server		
Effect Size	**Outcome**	**Cohen's *d***	**95% Confidence Interval**
	Based on PET-GAS change scores (unit = rating) [intent-to-treat analysis used]		
	PET-GAS (Δ from baseline) *Telepractice vs. On-site*	–0.30[b,c]	–0.98–0.39
	PET-GAS (Δ from baseline) *Telepractice vs. Placebo*	1.21	0.47–1.96

[a] included in combined effect

[b] included in meta-regression

[c] included in meta-analysis

Appendix B. Study Quality Rating Protocol

Quality Metric		Category	Score
Power		Inadequate	0
		Adequate	20
Design		Level 5	4
		Level 4	8
		Level 3	12
		Level 2	16
		Level 1	20
Internal Validity	Diagnosis	Different	0
		Same	2
	Group Matching	No statistical comparison, no matching reported	0
		No statistical comparison, matching reported	1
		Statistical comparison yields difference	2
		Statistical comparison yields similarity	3
	Treatment Protocol	Different	0
		Same	2
	Number of Sessions	Different	0
		Same	2
	Service Delivery Method	Different	0
		Same	2
	Treatment Targets	Different general targets	0
		Same general targets, different individual targets	1
		Same individual targets	2
	Outcome Measures	Not objective	0
		Objective	3
	Assessor Blinding	None	0
		Partial	2
		Total	4
External Validity	Study Population	Not applicable	0
		Applicable	10
	Service Delivery Model	Infeasible length/frequency	0
		Feasible length/frequency	5
		Infeasible service delivery method	0
		Feasible service delivery method	5

Appendix B. Study Quality Rating Protocol (continued)

Quality Metric		Category	Score
Reliability	Interrater Reliability	< 20% of outcome data	0
		≥ 20% of outcome data	5
		< 80% agreement	0
		≥ 80% agreement	5
	Treatment Integrity	< 20% of treatment sessions	0
		≥ 20% of treatment sessions	5
		< 80% correct implementation	0
		≥ 80% correct implementation	5
Highest Possible Quality Score			100