Volume 1 Number 1 April 2006

EBP briefs

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

Classroom-Based versus Pull-Out Interventions: A Review of the Experimental Evidence

Anita S. McGinty & Laura Justice
University of Virginia

A publication of AGS Publishing, now part of Pearson Assessments

EBP Briefs

A publication of AGS Publishing, now part of Pearson Assessements

Cite this document as:
McGinty, A. S., & Justice, L. M.
(2006). Classroom-based versus pull-
out speech-language intervention:
A review of the experimental evidence.
EBP Briefs, 1(1), 1-25.

EBP Briefs
Volume 1 Number 1 April 2006

Classroom-Based versus Pull-Out Interventions: A Review of the Experimental Evidence

Anita S. McGinty & Laura Justice
University of Virginia

EBP Briefs: An Introduction
Laura Justice, Editor

I am pleased to introduce an exciting and innovative venture supported by AGS Publishing, now part of Pearson Assessments: EBP Briefs. EBP Briefs come at an exciting time in our profession: As speech-language pathologists are well aware, there is currently a strong emphasis on the need to critically examine the best available current scientific evidence and integrate this evidence with other types of data to arrive at the best solutions to clinical questions. In this regard, clinical professionals are asked to bridge the "research to practice gap" by conducting reviews of the empirical literature in search of objective answers to questions faced in the field; subsequent to such review, they are expected to incorporate the resultant empirical evidence into their decision-making frameworks. This movement in speech-language pathology is not occurring in a vacuum; rather, it reflects a more general paradigm shift in the medical, allied health, and educational disciplines in which professionals are asked to consult the scientific evidence in their everyday decision-making, to ensure that theory and craft are balanced with data and evidence.

Evidence-based practice (EBP) is the term that describes the process clinical professionals go through as they consult various types of information to answer a clinical question. In the process of reviewing relevant information sources, the scientific literature is but one source consulted, but the tenets of EBP position this literature to be just as relevant as other information sources that are typically consulted (e.g., clinical experience, theoretical perspective). The clinician engaged in the process of EBP arrives at an answer to a clinical question that integrates information from an array of inputs, to include not only the scientific literature but also examination of client preferences, institutional norms and policies, and his/her theoretical knowledge and clinical experiences.

For clinicians to engage in the EBP process, they must have access to not only high-quality research studies relevant to the clinical questions they ask, but also to systematic reviews of particular bodies of evidence that attempt to find objective answers to commonly-asked questions. EBP Briefs is designed to support evidence-minded professionals by identifying and answering clinically-relevant questions using current standards of evidence-

based practice, with a particular focus on examining evidence relevant to pressing questions in school-based practice. Each brief considers a specific question, evaluates the available scientific evidence relevant to this question, and makes recommendations for integrating this scientific evidence into clinical decision-making.

EBP Briefs will be published in electronic format on a quarterly basis, and in a hard-copy journal format annually. Forthcoming briefs to look forward to in this 2006 inaugural volume include examinations of approaches used to address children's needs in social communication, vocabulary, fluency, and phonology. In planning for our second volume in 2007, I invite professionals to contact me to identify specific questions they encounter in clinical practice which we ought to systematically explore in EBP Briefs.

ljustice@virginia.edu

Classroom-Based versus Pull-Out Language Intervention: An Examination of the Experimental Evidence

Anita S. McGinty
Laura M. Justice
University of Virginia

In this brief, we consulted the available experimental evidence to consider an important question that clinicians often ask: Should I provide speech-language intervention within the child's classroom (classroom-based) or outside of the classroom (pull-out)? Following the tenets of evidence-based practice, we posed this question following the PICO framework (Centre for Evidence Based Medicine, 2001; Sackett, Straus, Richardson, Rosenberg, & Haynes, 1996), in which P represents the Patient, Patient Group, Population, or Problem; I represents the Intervention under consideration; C represents the Comparison intervention (i.e., either the standard approach or no intervention); and O represents the desired Outcome. Thus, our EBP question was this: Would a preschool or early-elementary child with language impairment (P) show greater improvement with classroom-based language intervention (individual or group) (I) or pull-out intervention (individual or group) (C), as shown by improvements in language skills in the areas of phonology, morphology/syntax, pragmatics, and/or vocabulary (O)?

This brief examines the research evidence supporting pull-out versus classroom-based models of language intervention.

Background and Rationale

When providing interventions to children with language impairment (LI) in the context of schooling, speech-language pathologists (SLPs) need to not only identify specific therapeutic targets and approaches, but also the intervention context or model of service delivery. Two general contexts/models prevail: pull-out and classroom-based. In pull-out models, the speech-language pathologist typically works with individual or small groups of children with communicative impairments in a workspace designated for this purpose.

Occasionally, peers without disabilities may participate in small-group activities to provide peer models during intervention. By contrast, classroom-based intervention is delivered within the classrooms of children with communicative impairments. Classroom-based models include several different approaches, typically differentiated into indirect and direct service delivery models based on the role of the SLP. With indirect services, the SLP serves as a consultant to the classroom teacher, providing expert guidance on adjusting targets and approaches to meet the needs of pupils with speech-language impairments within the classroom setting. With direct services, the SLP collaborates with the classroom teacher using a team-teaching approach (Meyer, 1997), or, alternatively, may deliver language lessons on his/her own to the class as a whole.

Relative to pull-out models, classroom-based interventions are often viewed as adhering to a more inclusive philosophy of education for children with disabilities. Inclusion is a philosophy that advocates education of pupils with disabilities in the same contexts of non-disabled peers (Meyer, 1997). By contrast, pull-out interventions are seen as non-inclusive, in that children with disabilities are removed from general education to receive intervention away from their peers. The possible value of classroom-based models of intervention goes beyond philosophy, however, and are viewed by some experts as likely more effective than pull-out models, particularly in the area of skill generalization. Classroom-based approaches may enhance skill generalization through their emphasis on naturalistic routines and materials, the involvement of peers as both conversational models and partners, and the involvement of teachers who can extend language instruction throughout the day (Wilcox, Kouri, & Caswell, 1991).

Historically, speech-language interventions often used pull-out models featuring individual or small-group treatment. Since 1975, however, changes in federal law as well as general philosophies towards the education and treatment of children with disabilities has motivated SLPs to consider more inclusive approaches to intervention (Ehren, 2000; Elksnin & Capilouto, 1994; Kavale, 2002; Manset & Semmel, 1997), including delivery of interventions directly within the general education classroom (Beck & Dennis, 1997; Ehren, 2000;

Westby, 1994). As innocuous as classroom-based models of service delivery may seem, use of this service delivery model is not without controversy (Beck & Dennis, 1997; Ehren, 2000; Kavale, 2002; Westby, 1994); researchers have reported concerns about the overall effectiveness of classroom-based services (Ehren, 2000; Elksnin & Capilouto, 1994), challenges in collaborative teaming (Achilles, Yates, & Freese, 1991; Beck & Dennis, 1997; Russell & Kaderavek, 1993) and concerns regarding the unilateral application of inclusive philosophies without regard to individual differences (Westby, 1994). Such concerns were summarized in a statement by the American Speech and Hearing Association (ASHA) warning that "the shift toward inclusion will not be optimal when implemented in absolute terms...the unique and specific needs of each child and family must always be considered" (1996, p. 1).

Within an EBP framework, clinicians make decisions about the context in which to provide school-based language interventions following the same process used when making choices related to other aspects of intervention programming, such as goal selection and sequencing, progress monitoring, and length and intensity of therapy. In making decisions regarding treatment contexts, SLPs who adhere to the tenets of EBP consider their theories and experiences with various contexts of service delivery, the specific needs of children and families, as well as the overall culture and norms of the schools in which they work; additionally, they also look to the accumulated scientific literature for objective guidance concerning the effectiveness of treatment within different contexts (i.e., classroom-based vs. pull-out).

Evidence-based practice requires SLPs to consider their theories, experiences, the specific needs of clients, and the accumulated scientific literature when making clinical decisions

Of particular relevance to the SLP considering which service delivery context to use with a particular child are experimental studies that have compared treatment outcomes for classroom-based versus pull-out models. Although a variety of research designs can provide important evidence to draw upon in clinical decision-making (e.g., see Horner, Carr, Halle, McGee, Odom, & Wolery, 2005; Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005), the most rigorous design is that of the experimental study, which is designed specifically to establish causality (Law, Garrett, & Nye, 2005). Accordingly,

many organizations that oversee development of systematic reviews that synthesize the evidence on a particular intervention privilege experimental studies (particularly randomized clinical trials) as providing the strongest type of evidence, when the study is of sufficient quality (What Works Clearinghouse, 2005).

A number of experimental or quasi-experimental studies have directly investigated the impact of classroom-based interventions on children's language outcomes, but relatively few of these have directly compared children's outcomes for classroom-based models relative to pull-out models of intervention. In this review, in which our goal was to synthesize the available scientific evidence relevant to the question of whether clinicians ought to use classroom-based or pull-out language interventions, we focused specifically on those studies that directly compared these two models of service delivery.

Method

Inclusionary/Exclusionary Criteria

Prior to searching for studies relevant to our clinical question, we developed a set of inclusionary/exclusionary criteria (see Table 1) for identifying studies to be included in this review. Our first criterion was that the study featured an experimental design (randomized clinical trial; RCT), quasi-experimental design (QED), or regression discontinuity design (RD), as these three types of designs provide the best evidence of causality (Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti, 2005). The U.S. Department of Education's What Works Clearinghouse (www.whatworks.ed.gov), which publishes systematic reviews of research to identify educational interventions with sufficient evidence to support their use, considers only evidence from studies using these three designs in their reviews. In these designs, the focus of investigation is estimating the causal relationship between a set of independent variable(s) and a set of dependent/outcome variable(s). To be included in the present review, the independent variable(s) needed to represent a comparison of classroom-based versus pull-out models of language intervention, whereas the dependent measures needed to represent child language outcomes in expressive and/or receptive language (phonology, vocabulary, syntax/morphology, and/or

pragmatics).

Additional criteria for inclusion of studies specified participants' age range (two to eight years) and type of disability; for the latter, we included studies involving children with specific language impairment (SLI; receptive and/or expressive) and excluded studies involving children with language impairment secondary to cognitive and sensory disabilities (e.g., mental retardation, autism), similar to Law et al. (2004). As much as 6 to 8% of young children exhibit SLI (Tomblin et al., 1997), and with its clear impact on both early and later academic achievement (Aram, Ekelman, & Nation, 1984; Scarborough & Dobrich, 1990), it is one of the most common disabilities addressed by special educators working in the public schools. Final criteria for study inclusion concerned publication venue, to include only English-language studies published in peer-reviewed journals.

Article Search

A comprehensive search protocol was developed to identify studies meeting the inclusionary criteria identified previously. Databases searched included PsychINFO; Educational Resources Information Center (ERIC); Campbell Collaboration Social, Psychological, Education, and Criminological Trials Register; Cochrane Controlled Trials Register; American Speech-Language-Hearing Association Journal Archives; What Works Clearinghouse; and the Scottish Intercollegiate Guidelines Network. Additionally, Law and colleague's (2004) recent meta-analysis of speech-language treatment efficacy was also examined. A total of 783 papers were screened for inclusion, and of these, 34 were identified for further examination of the full-text article. Of these 34 articles, 23 were eliminated because they did not use an RCT, RD, or QED design; five because they did not directly compare classroom-based vs. pull-out intervention; and three because the study sample did not reflect our inclusionary criteria. Therefore, the final corpus available for this review was three studies. This figure, while low, is not surprising given that Law and colleagues (2004) found only 13 studies for their systematic review of treatment outcomes for children with SLI receiving any type of language

Three experimental studies directly compared child outcomes in classroom-based versus pull-out models of language intervention.

intervention.

In Appendix 1, we provide a description of articles that surfaced in our review that provided comparisons of various service delivery models and contexts for children with a range of abilities/disabilities; readers can consult these works for additional guidance in their evidence-based decision-making concerning treatment contexts. Table 2 presents a summary of the three studies included in the present review.

Results

Description of Included Studies

The three studies included in this review (see Table 2) involved a total sample size of 91 children with SLI: 20 between the ages of 20 to 47 months (Wilcox et al., 1991), 39 between 3 and 5 years (Valdez & Montgomery, 1997), and 31 in kindergarten through third grade (Throneburg, Calvert, Sturm, Paramboukas, & Paul, 2000). Fifty-three children received language intervention in a classroom-based inclusive setting using a team-teaching collaborative model (n = 42) or SLP-lead inclusive instruction with the SLP working independently of the classroom teacher (n = 11). Thirty-eight children received pull-out intervention in individual or small-group sessions. The duration of intervention ranged from three months to six months, and children typically received one or two treatment sessions weekly. Outcome measures in two studies focused specifically on vocabulary, documenting children's growth in vocabulary using an experimenter-developed test (Throneburg et al., 2000) or language sample analysis (Wilcox et al., 1991); one study used a norm-referenced standardized test of language as the primary outcome measure (Valdez & Montgomery, 1997).

Quality of Evidence

As Law and colleagues (2004) point out, experimental studies vary tremendously in the quality of their design. For descriptive purposes, we conducted a quality assessment of the three studies included in this review, focusing on 11 attributes of high-quality studies using Law et al.'s 3-point scale (0=inadequate, 1=unclear, 2=adequate), as presented in Table 3. Appendix

2 provides a description of each attribute evaluated; the 11 attributes were selected from available descriptions of high-quality indicators of experimental and quasi-experimental research. Each of the authors independently scored each of the three studies for these 11 attributes; scores were compared to establish inter-rater reliability of the scoring protocol, which was 90%.

There is currently little agreement on what differentiates higher and lower quality studies, although studies of higher quality provide the most rigorous test of causality for the intervention under investigation. The What Works Clearinghouse standards of evidence for establishing causality of an intervention require RCTs to use randomization, ensure baseline equivalence, and address attrition, each of which was rated as adequate for the two RCTs included in this review. Quasi-experimental designs that ensure baseline equivalence and address attrition can meet evidence standards, albeit with reservations given that causality is less certain due to lack of randomization. The single QED included in this review was rated as adequate on these attributes. Thus, in terms of design quality, the three studies included in this review were of sufficient quality to establish causal relationships between intervention contexts (classroom-based vs. pull-out) and children's language outcomes.

Estimated Effects: Classroom-Based and Pull-Out Intervention

Table 4 summarizes study outcomes based on the authors' descriptions; we also present our own calculation of effect-size estimates (standardized posttest differences, corrected for bias using Hedges and Olkin's factor, 1985) with 95% confidence intervals based on posttest data from the study reports. Per Cohen (1988), we considered an effect-size estimate (d) of 0.2 as small, 0.5 as medium, and 0.8 as large. Effect-size estimates represent the difference in scores between two groups using standard deviation units; thus, an effect-size of d = 0.95 calculated on the posttest scores between two groups suggests that one group's posttest scores were nearly one standard deviation higher than those of the other group, consistent with a large intervention effect. Note that effect size estimates could not be calculated for Valdez and Montgomery (1997), as the report did not include standard deviations for children's posttest data. Considering both the authors' own analyses of child outcomes and our

calculations of effect-size estimates on posttest data, two studies (Throneburg et al., 2000; Wilcox et al., 1991) show better outcomes for children in classroom-based interventions that involve team-teaching in inclusive classrooms (i.e., classrooms containing peers without disabilities), whereas one study reports no differences in child outcomes when comparing classroom-based and pull-out conditions (Valdez & Montgomery, 1997).

Examining the findings from the three studies more closely, Throneburg et al. (2000) reported that children in the collaboration condition, in which SLPs team-taught language lessons with the classroom teachers, exhibited greater gains in curricular vocabulary compared to children in a classroom-based condition in which SLPs taught alone (d = 1.56, CI = 0.63-2.5) and children in a pull-out condition (d = 0.31, CI = -0.56-1.18). The difference between posttest vocabulary scores for children in the team-taught classroom-based intervention and the SLP-alone classroom-based intervention was very large when considering effect-size estimates. And, although the difference in curricular vocabulary outcomes for children in the classroom-based and pull-out conditions was not statistically significant, based on the authors' inferential test statistics, our calculation of effect-size estimates suggested a clear advantage for pull-out over classroom-based intervention in which the SLP taught alone (d = -0.76, CI = -1.67-0.15).

Throneburg and colleagues found classroom-based team-teaching to be more effective for improving children's curricular vocabulary knowledge than pull-out intervention.

Like Throneburg and colleagues, Wilcox et al. (1991) found a significant difference in children's productive vocabulary use, particularly when considering children's vocabulary use in the home setting, for classroom-based team-teaching (SLP + special educator) compared to a pull-out condition (d = 0.81, CI = -0.1-1.72). Again, the effect-size estimate on posttest scores was consistent with a large effect.

Valdez and Montgomery's findings diverged from the previous two studies reviewed, showing no difference in classroom-based and pull-out conditions for children's total language and expressive language scores but reporting an advantage for the pull-out condition on children's receptive language. It is difficult to assess, however, the strength of causal interpretations

for Valdez and Montgomery (1997), as the report does not provide test statistics for comparisons of children's performance at posttest, nor are standard deviations included in their descriptive tables for our own calculation of posttest effect-size estimates. We therefore were unable to calculate effect-size estimates on posttest differences between the two groups. As a result, we considered Throneburg et al. (2000) and Wilcox et al. (1991) to provide the strongest causal interpretations regarding child language outcomes in the two service delivery contexts (classroom-based vs. pull-out): both showed an advantage for vocabulary development in classroom-based inclusive team-teaching approaches compared to pull-out instruction, with effect-size estimates consistent with large to very large differences favoring classroom-based models.

Conclusions

Would a preschool or early-elementary child with language impairment (P) show greater improvement with classroom-based language intervention (individual or group) (I) or pull-out intervention (individual or group) (C), as shown by improvements in language skills in the areas of phonology, morphology/syntax, pragmatics, and or vocabulary (O)?

Scientific inquiry is, ultimately, about accumulating evidence and establishing convergence of findings across studies. As clinical professionals, we ought not to change practices in light of evidence from one study; nonetheless, when several well-conducted studies converge in their findings, it does suggest that evidence from these studies ought to warrant careful consideration in evidence-based decision-making.

As we reported in this review of three studies comparing pull-out and classroom-based models of language intervention, two of three studies converged in their findings of a benefit to collaborative classroom-based services over pull-out services for preschool and early elementary pupils when addressing vocabulary goals. In interpreting the divergent findings from the third study, Valdez and Montgomery (1997), considerations regarding the model of classroom-based intervention and the means of measuring outcomes provide possible explanations for their difference in findings. Throneburg et al. (2000) and Wilcox et al. (1991) used a team-teaching approach in their

classroom-based models, in which the SLP and special educator worked side-by-side; in Valdez and Montgomery (1997), however, the description of the classroom-based condition was not sufficient for determining if the classroom-based approach involved a team teaching approach or SLP-lead instruction occurring separately from teacher-lead instruction. As Throneburg et al. (2000) showed, these two types of classroom-based service delivery models differed in their effects on children's language outcomes; in fact, Throneburg et al. found SLP-lead lessons (occurring in absence of the classroom teacher) was less effective than pull-out models and a team-teaching classroom-based approach. An additional point for consideration is that Valdez and Montgomery's (1997) outcome measure was a standardized test of general language, which may not have been well aligned or sensitive to changes in children's language achievements during intervention; by comparison, both Throneburg et al. (2000) and Wilcox et al. (1991) used outcome measures carefully aligned to measuring aspects of language targeted in intervention. Taken together, the convergent findings from Throneburg et al. (2000) and Wilcox et al. (1991) suggested an advantage for classroom-based team-teaching models over pull-out intervention, at least in the domain of vocabulary.

Limitations in the number of studies included in this review, the strength of effect-size estimates, and weaknesses in study quality require that these review findings are interpreted cautiously. The fact that only three studies were able to meet design-based evidence standards for inclusion in this review highlights the clear dearth of evidence regarding various service-delivery models, and the lack of a clear effect for classroom-based over pull-out intervention in one of the three studies suggests the need for future, rigorous evaluations of classroom-based models of therapy for more conclusive answers. Additionally, while medium to large effect sizes were found in Throneburg et al. (2000) and Wilcox et al. (1991) for collaborative team-teaching in comparison to pull-out therapy, the 95% confidence intervals for some estimates included 0, raising the possibility that neither approach (classroom-based vs. pull-out) provided an advantage in children's vocabulary outcomes. Finally, weaknesses in the studies pose potential threats to the strengths of the studies' findings. None of the studies used assessors who were blind to treatment conditions, which

potentially allows for a bias on the part of the assessor to impact results. Only Wilcox et al. (1991) provided information on fidelity to treatment and only Valdez and Montgomery (1997) provided sufficient information regarding the use of equivalent service providers across conditions, raising questions as to whether outcomes were also impacted by the process of service delivery and/or characteristics of the treatment providers. Future studies that compare various service-delivery options must include more rigorous design features, including blind assessors. As a final limitation, we also note that with only three studies serving as our corpus for review, it is not clear that results can be generalized. As each of the three studies involved children from a single geographic region, from primarily one race (either Caucasian or African American), and from lower to middle socioeconomic backgrounds, large scale effectiveness studies should investigate the impact of these treatment models on socially, ethnically and geographically diverse population to improve generalization of findings.

In sum, this review considered three experimental studies comparing child language outcomes for two prevalent service-delivery contexts: classroom-based and pull-out. Systematic review of the outcomes showed an advantage for classroom-based inclusive models in which the SLP and classroom teacher team-taught language lessons. With the framework of evidence-based practice, clinicians can integrate the evidence presented in this review with other sources of information, such as child and family preferences, their own experiences with various models of service delivery, and the culture in which they work, to make the best decisions concerning the models of service delivery they use to meet the needs of the children with whom they work.

References

Achilles, J., Yates, R. R., & Freese, J. M. (1991). Perspectives from the field: Collaborative consultation in the speech and language program of the Dallas Independent School District. *Language, Speech, and Hearing Services in Schools, 22,* 154-155.

American Speech-Language-Hearing Association (1996, Spring). Inclusive practices for children and youths with communication disorders: Position statement and technical report. *Asha, 38* (Suppl. 16), 35-44.

Aram, D. M., Ekelman, B. L., & Nation, J. E. (1984). Preschoolers with language disorders: 10 years later. J*ournal of Speech and Hearing Research, 27,* 232-244.

Beck, A. R., & Dennis, M. (1997). Speech-language pathologists' and teachers' perceptions of classroom-based interventions. *Language, Speech, and Hearing Services in Schools, 28,* 146-153.

Centre for Evidence Based Medicine (2001). *Levels of evidence and grades of recommendation.* Available at http://www.cebm.net/levels_of_evidence.asp.

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

Cole, K. N., Mills, P. E., Dale, P. S., & Jenkins, J. R. (1996). Preschool language facilitation methods and child characteristics. *Journal of Early Intervention, 20(2),* 113-131.

Ehren, B. J. (2000). Maintaining a therapeutic focus and sharing responsibility for student success: Keys to in-classroom speech-language services. *Language, Speech, and Hearing Services in Schools, 31,* 219-229.

Elksnin, L. K., & Capilouto, G. J. (1994). Speech-language pathologists' perceptions of integrated service delivery in school settings. *Language, Speech, and Hearing Services in Schools, 25,* 258-267.

Ellis, L., Schlaudecker, C., & Regimbal, C. (1995). Effectiveness of a collaborative consultation approach to basic concept instruction with kindergarten children. *Language, Speech, and Hearing Services in Schools, 26,* 69-73.

Farber, B. A., & Klein, E. R. (1999). Classroom-based assessment of a collaborative intervention program with kindergarten and first-grade students. *Language, Speech, and Hearing Services in Schools, 30,* 83-91.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research. *Exceptional Children, 71,* 137-149.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

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

Kavale, K. A. (2002). Mainstreaming to full inclusion: from orthogenesis to pathogenesis of an idea. *International Journal of Disability, Development and Education, 49(2),* 201-214.

Law, J., Garrett, Z., & Nye, C. (2005). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 47,* 924-943.

Manset, G., & Semmel, M. I. (1997). Are inclusive programs for students with mild disabilities effective? A comparative review of model programs. *The Journal of Special Education, 31,* 155-180.

Meyer, J. (1997). Models of service delivery. In P. O'Connel (Ed.), *Speech, language, and hearing programs in schools: A guide for students and practitioners.* Gaithersburg, MD: Aspen Publishers.

Rafferty, Y., Piscitelli, V., & Boettcher, C. (2003). The impact of inclusion on language development and social competence among preschoolers with disabilities. *Exceptional Children, 4,* 467-479.

Roberts, J. E., Prizant, B., & McWilliam, R. A. (1995). Out-of class versus in-class service delivery in language intervention: Effects on communication interactions with young children. *American Journal of Speech-Language Pathology, 69(4),* 87-94.

Russell, S. C., & Kaderavek, J. (1993). Alternative models for collaboration. *Language, Speech, and Hearing Services in Schools, 24,* 76-78.

Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W., & Haynes, R.B. (2000). *Evidence-based medicine: How to practice and teach EBM.* Edinburgh: Churchill Livingstone.

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech and Hearing Research, 33,* 70-83.

Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71,* 181-194.

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. A., & Paul, P. J. (2000). A comparison of service delivery models: Effects of curricular vocabulary skills in the school setting. *American Journal of Speech-Language Pathology, 9,* 10-20.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245-1260.

Valdez, F. M., & Montgomery, J. K. (1997). Outcomes from two treatment approaches for children with communication disorders in Head Start. *Journal of Children's Communication Development, 18,* 65-71.

Westby, C. E. (1994). The vision of full inclusion: Don't exclude kids by including them. *Journal of Childhood Communication Disorders, 16,* 13-22.

What Works Clearinghouse (2005). *WWC study review standards.* Available at: http://www.whatworks.ed.gov/reviewprocess/study_standards_final.pdf.

Wilcox, M. J., Kouri, T. A., & Caswell, S. B. (1991). Early language intervention: A comparison of classroom and individual treatment. *American Journal of Speech-Language Pathology, 1,* 49-60.

Yoder, P. J., Kaiser, A. P., Goldstein, H., Alpert, C., Mousetis, L., Kaczmarek, L., et al. (1995). An exploratory comparison of milieu teaching and responsive interaction in classroom applications. *Journal of Early Intervention, 19(3),* 218-242.

Table 1. Inclusionary and Exclusionary Criteria for Studies

Design
Randomized Controlled Trial (RCT) OR Quasi-experimental design (QED) OR Regression discontinuity design (RD)
Independent Variable(s)
Direct comparison of classroom-based and pull-out services AND Intervention targets expressive and/or receptive language skills in phonology, vocabulary, syntax/morphology, and/or pragmatics
Dependent Variable(s)
Direct measures of receptive and/or language skills in phonology, vocabulary, syntax/morphology, and/or pragmatics
Participants
Children between 2 and 8 years of age AND Primary impairment of language (receptive and/or expressive impairment in phonology, vocabulary, syntax/morphology, and/or pragmatics)
Publication
Peer-reviewed journal AND English language

Table 2. Description of Studies Comparing Classroom-Based and Pull-Out Language Intervention.

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. A., & Paul, P. J. (2000). A comparison of service delivery models: Effects of curricular vocabulary skills in the school setting. American Journal of Speech-Language Pathology, 9, 10-20.	
Participants	31 children (primarily Caucasian) in kindergarten through third grade attending 12 classrooms in two elementary schools. An additional 146 typically-developing peers in their classrooms participated. Children identified with speech and language impairment based on standardized testing by school SLPs. Performance of -1SD of the mean or lower on two tests required for eligibility.
Research Design	Quasi-Experimental Design using 12 intact classrooms in three schools assigned to three conditions (4 classrooms per condition). Eight of 12 classrooms randomly assigned to two conditions (classroom-based, pull-out); remaining four classrooms assigned to collaborative condition for administrative reasons.
Experimental Intervention 1	"Collaborative:" Collaborative team-taught language lessons to whole class by SLP, classroom teacher, and two graduate students in weekly 40-min sessions for a 12-week period (8 hours of treatment total) combined within an additional 15-min pull-out small-group session.
Experimental Intervention 2	"Classroom-Based:" SLP-delivered language lessons to whole class in weekly 40-min sessions for a 12-week period (8 hours of treatment total) combined within an additional 15-min pull-out small-group session.
Comparison Intervention	"Pull-Out:" Weekly 50-min small-group or individual pull-out sessions held in speech room for 12 weeks.
Outcome Measure	Researcher-developed vocabulary test of curricular vocabulary (one for each grade level) comprising 20 words, with up to 60 points possible.
Valdez, F. M., & Montgomery, J. K. (1997). Outcomes from two treatment approaches for children with communication disorders in Head Start. Journal of Children's Communication Development, 18, 65-71.	
Participants	39 African-American children ranging from 3 to 5 years (M = 3.9 years) enrolled in four Head Start centers. Children identified with SLI using standardized test of language by certified SLPs. (Exact criteria not specified.)
Research Design	Randomized Clinical Trial comparing two conditions using a randomized block design with stratification by severity of language impairment (mild, moderate, severe).

Table 2, Continued

Experimental Intervention	"Inclusion Group:" 90-min treatment session weekly for six months (36 hours of treatment total) conducted in Head Start classroom with 10 to 15 typical peers. Specific approach and targets not identified.
Comparison Intervention	"Pull-Out Group:" 90-min treatment session weekly for six months (36 hours of treatment total) conducted in small groups in a "small separate room" at a Head Start center. Neither the size of groups nor specific approaches and targets were identified.
Outcome Measure	Standardized test scores on six subtests of the Clinical Evaluation of Language Fundamentals-Preschool (CELF-P; Wiig, Secord, & Semel, 1991): Linguistic Concepts, Basic Concepts, Sentence Structure, Recalling Sentences, Formulating Labels, and Word Structure. Total Language Score, Receptive Language Score, and Expressive Language Score also reported.
Wilcox, M. J., Kouri, T. A., & Caswell, S. B. (1991). Early language intervention: A comparison of classroom and individual treatment. American Journal of Speech-Language Pathology, 1, 49-60.	
Participants	20 children (race/ethnicity unspecified) ranging from 20 to 47 months (M = 3.9 years). Children identified with SLI using standardized test of language conducted at a university speech and hearing clinic; performance of -1.5 SD of the mean or lower required for eligibility.
Research Design	Randomized Clinical Trial comparing two conditions.
Experimental Intervention	"Classroom Intervention:" Twice-weekly treatment sessions for 12-16 weeks conducted in classroom of 12 to 14 children. Intervention delivered by team (SLP and early childhood special educator) using interactive modeling of vocabulary targets during regular classroom activities across a 3-hour period.
Comparison Intervention	"Individual Intervention:" Twice-weekly 45-min one-on-one treatment sessions for 12-16 weeks in speech/language room. Intervention delivered by student SLP using interactive modeling of vocabulary targets.
Outcome Measure	Number of words used productively in spontaneous speech (i.e., with phonetically consistent adult-like form and used in at least three different contexts) measured in 200-utterance language sample.

Table 3. Evaluation of Study Quality based on Law, Garrett, and Nye's (2004) 3-Point Scale (0=inadequate; 1=unclear; 2=adequate)

	Study		
	Valdez & Montgomery (1997)	Throneburg et al. (2000)	Wilcox et al. (1991)
Criteria[1]			
Randomization[2]	2	0	2
Baseline Equivalence	2	2	2
Attrition	2	2	2
Participant Description	2	2	2
Comparable Trainers	2	1	1
Treatment Detail	0	2	2
Treatment Fidelity	0	0	2
Blinding of Assessors	1	1	1
Reliable Outcomes Measures	1	1	1
Generalized Performance	0	0	2
Effect Size Estimates[3]	0	2	0

1 The grey rows indicate qualities that the What Works Clearinghouse (U. S. Department of Education, 2005) considers necessary for a study to provide strong evidence of causality.
2 Randomization occurred for two of three conditions (classroom-based and pull-out but not for collaborative condition.)
3 Valdez & Montgomery do not provide effect-size estimates or standard deviations for outcome measures, thus effect-size estimates cannot be independently calculated; Throneburg et al. provide an omnibus effect-size estimate but none for univariate comparisons, although the appropriate data are included for independent calculation; Wilcox et al. do not provide effect-size estimates, but the appropriate data are included for independent calculation.

Table 4. Description of Child-Language Outcomes for Three Studies

Study (Treatment Comparisons)	Outcome Measure(s)	Study's Description of Findings	Standardized posttest difference (95% confidence interval)
Throneburg et al. (2000) (Collaboration/ Team-Teaching, Classroom-Based, Pull-Out)	Curricular vocabulary	Children in the collaboration condition made significantly greater gains than children in the classroom-based and pull-out conditions, who performed similarly at posttest	Collaboration vs. classroom-based: d = 1.56 (0.63 to 2.5) Collaboration vs. pull-out: d = 0.31 (-0.56 to 1.18) Classroom-based vs. pull-out: d = -0.76 (-1.67 to 0.15)
Valdez & Montgomery (1996) (Inclusion, Pull-Out)	Expressive and receptive language	Children in the pull-out group made greater gains in receptive language scores than children in the inclusion group (test statistic not reported); performance was similar across both groups on expressive and total language scores.	Standard deviations not provided, thus standardized posttest differences could not be calculated.
Wilcox et al. (1991) (Classroom, Intervention)	Production of target vocabulary	Children in the classroom condition used significantly more target words when averaged for home and treatment settings	Classroom vs. individual for vocabulary in treatment + home setting: d = 0.63 (-0.27 to 1.53) Classroom vs. individual for vocabulary in treatment setting: d = 0.2 (-0.68 to 1.08) Classroom vs. individual for vocabulary in home setting: d = 0.81 (-0.1 to 1.72)

Appendix 1

Reference	Design	Setting and Population Characteristics	Intervention/ Comparison	Characteristics of Intervention(s)	Outcome/Results
Cole, Mills, Dale & Jenkins, 1996	Randomized clinical trial N=6 classrooms; 52 children	Reverse inclusion; study children had significant language delays (10 months or greater) Mean age= 52 months	Direct Language (DirL) vs Developmental Language (DevL) Programs	Direct Classroom based in small/large group Parallel teach	Standardized language tests in across receptive/expressive language areas No differential gains seen across treatments
Ellis, Schlaudecker, & Regimbal, 1995	Quasi-experimental, pre-post control group N= 2 classrooms, 40 children	General education kindergarten classrooms in demographically at risk area Mean age= 5.6	Collaborative consultation vs. regular curriculum (no intervention)	Indirect; Collaborative consultation Classroom based	Basic concepts Significant growth in collaborative classroom
Farber & Klein, 1999	Quasi-experimental with random assignment at classroom level. Matched pairs of children in experimental and control classrooms N=6 schools; 12 classrooms; 522 children	General education kindergarten and first grade classrooms Heterogeneous populations in terms of abilities and demographics	Collaborative classroom vs. inclusion with no collaboration with primary teacher vs. no intervention	Indirect Classroom based	Receptive/expressive oral language; reading; writing Children in collaborative classroom or receiving classroom based services (out of primary classroom) were significantly better in receptive language and writing
Rafferty, Piscitelli, & Boettcher, 2003	Pre-post control group, no randomization N= 96 children (number of classrooms not specified)	Preschool classrooms serving children with disabilities that were inclusive or self-contained; children in the study had a range of mild to severe developmental delays	Parallel teach (special educator and early childhood educator) vs. no collaboration	Direct, Collaborative teaching Classroom based	Language and social skills, and behavior No significant difference between inclusive classrooms or self-contained classrooms in language or social skills growth
Roberts, Prizant, & McWilliam, 1995	Randomized clinical trial to treatment conditions N= 15	Mainstreamed daycare; study children had mild to moderate developmental delays Mean age= 34 months	Inclusion vs. pull-out	Direct Classroom or pull-out	SLP and child turn-taking; child communication intents and responsiveness Fewer turns by SLPs and fewer responses by children in classroom sessions

Appendix 1

Reference	Design	Setting and Population Characteristics	Intervention/ Comparison	Characteristics of Intervention(s)	Outcome/Results
Yoder, Kaiser, Goldstein, Alpert, Mousetis, Kaczmarek & Fischer, 1995	Quasi-experimental pre-post control group N=6 classrooms; 36 children	ECSE classrooms serving children with mild to severe developmental delays; study children had moderate to severe language delays Age range= 2-7 years	Milieu Teaching (MT) vs. Responsive Interaction (RI)	Indirect; Collaborative consultation Classroom based	Vocabulary MT more effective than RI for children with pretest language scores at or below 33-34 months RI more effective for children with pretest language scores at or above 41-45 months

Appendix 2
Indicators of High-Quality Experimental and Quasi-Experimental Studies

Randomization	Participants are randomly assigned to conditions.
Participant Description	Characteristics of participants are adequately described, including SES, race/ethnicity, and age; for studies of children with disabilities, procedures for confirming/determining presence of disability are included.
Baseline Equivalence	Participants in each condition are equivalent at baseline on relevant characteristics, or lack of equivalence addressed using appropriate techniques.
Comparable Trainers	Procedures used to ensure that trainers/interventionists are comparable across conditions.
Attrition	Participant attrition from conditions is comparable and accounted for in analyses.
Treatment Detail	Treatment described in sufficient detail to permit its replication.
Treatment Fidelity	Procedures used to ensure treatment implemented in the way described.
Blinding of Assessors	Assessors blind to participants' treatment conditions.
Reliable Outcome Measures	Reliability estimates or evidence provided for key outcome measures.
Generalized Performance	Outcome measures include measures closely aligned to intervention as well as measures of generalized performance.
Effect Size Estimates	Effect-size estimates provided and interpreted.

Sources: Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti (2005); What Works Clearinghouse (2005)

EBP Briefs
Volume 1 Number 1 April 2006

Classroom-Based versus Pull-Out Interventions: A Review of the Experimental Evidence

Anita S. McGinty & Laura Justice
University of Virginia

A publication of AGS Publishing, now part of Pearson Assessments