

# Twenty Years of Communication Intervention Research With Individuals Who Have Severe Intellectual and Developmental Disabilities

**Martha E. Snell**  
University of Virginia

**Nancy Brady**  
University of Kansas

**Lee McLean**  
University of North Carolina at Chapel-Hill

**Billy T. Ogletree**  
Western Carolina University

**Ellin Siegel**  
University of Nebraska

**Lorraine Sylvester**  
University of Oklahoma

**Beth Mineo**  
University of Delaware

**Diane Paul**  
American Speech-Language-Hearing Association (Rockville, MD)

**Mary Ann Ronski and Rose Sevcik**  
Georgia State University

## Abstract

This literature review was conducted to evaluate the current state of evidence supporting communication interventions for individuals with severe intellectual and developmental disabilities. We reviewed 116 articles published between 1987 and 2007 in refereed journals meeting three criteria: (a) described a communication intervention, (b) involved one or more participants with severe intellectual and developmental disabilities, and (c) addressed one or more areas of communication performance. Many researchers failed to report treatment fidelity or to assess basic aspects of intervention effects, including generalization, maintenance, and social validity. The evidence reviewed indicates that 96% of the studies reported positive changes in some aspects of communication. These findings support the provision of communication intervention to persons with severe intellectual and developmental disabilities. Gaps in the research were reported as were recommendations for future research.

DOI: 10.1352/1944-7558-115-5.364

The ability to communicate effectively with others is essential for good quality of life. Individuals who have severe disabilities include those with severe to profound intellectual disability, autism, deaf-blindness, and multiple disabilities. For these individuals, the ability to communicate can be substantially compromised. The question of whether and how this ability to communicate can be improved through intervention was the focus of a national consensus conference convened by the U.S. Department of Education's Office of Special Education Programs (OSEP) and its Technical Assistance Development System (TADS) in 1985 (Office of Special Education, 1985). In addition to producing a number of consensus statements, these 1985 conferees called for the formation of an interagency task force to disseminate guidelines for the development and enhancement of functional communication abilities in individuals with severe disabilities.

This recommendation resulted in the establishment of the National Joint Committee for the Communicative Needs of Persons With Severe Disabilities (NJC) in 1986. The present review was conducted by current members of the NJC, which included representatives from the American Association of Intellectual and Developmental Disabilities, American Occupational Therapy Association, American Physical Therapy Association, American Speech-Language-Hearing Association, Council for Exceptional Children/Division for Children With Communication Disabilities and Deafness, TASH (formerly The Association for Persons With Severe Handicaps), and the United States Society for Augmentative and Alternative Communication.

### *Evidence-Based Practice*

In the past 5 years, much has been written about the importance of basing medical, therapeutic, and educational interventions on high quality empirical evidence. This focus on the need for more evidence-based practice can be found across all disciplines represented on the NJC in the form of articles, position statements, and special issues of our journals. Although there is no universal agreement on what constitutes evidence-based practice or how to evaluate the relative quality of available evidence, there are some clear areas of agreement. Echoing the terminology introduced by Sackett and colleagues (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996),

the American Speech-Language-Hearing Association issued a position statement that noted "the term 'evidence-based practice' refers to an approach in which current, high-quality research evidence is integrated with practitioner expertise and client preferences and values into the process of making clinical decisions" (American Speech-Language-Hearing Association, 2005, p. 1). Similar statements and definitions have been issued by the American Psychological Association (2005), the American Occupational Therapy Association (Guttman, 2009), the American Physical Therapy Association (n.d.), and the Council for Exceptional Children (Odom et al., 2005).

The multiple professions represented on the NJC also share some common expectations as to what constitutes high quality research evidence. It is generally accepted that the highest level of evidence quality is produced by a *randomized clinical trial*, a prospective study in which researchers use randomized assignment of participants with double-blind controls (Committee on Educational Interventions, 2001; Shavelson & Towne, 2002). However, in research on intervention provided to individuals with severe intellectual and developmental disabilities, often compounded by multiple physical and sensory disabilities, such designs are usually not possible for a number of practical, ethical, and scientific reasons. (Scientific standards for a true randomized clinical trial require assumptions about equal distribution of traits and representativeness of population samples that are not met in studies of individuals with very low incidence disabilities.) Thus, any attempt to evaluate the quality of evidence supporting different types of intervention for individuals with severe intellectual and developmental disabilities must consider the research characteristics that contribute to the overall believability of the research results, regardless of research design (Odom et al., 2005).

Most basically, results are more believable to the extent that a study design has controlled for threats to both *internal validity* (i.e., are results actually attributable to the experimental procedures, as described?) and *external validity* (i.e., are results useful and generalizable to other members of the target population?) (Tuckman, 1999). So, although it may not be realistic to look for randomized clinical trial designs in intervention studies involving individuals with severe intellectual and developmental disabilities, the design features themselves that contribute to internal and

external validity of these studies can be examined. These quality indicators are (a) accurate and complete description of participant characteristics, especially traits likely to be related to the dependent measures of the study; (b) replicability of study procedures, including precise description of how the procedure was implemented, the intensity (how often? for how long?), and duration of treatment (how many days/weeks/months?) and how reliably these procedures were implemented as described (*treatment fidelity*); (c) reliability of data reported (i.e., do these data accurately reflect participant characteristics and results of intervention?), including referring to both inter- and intrarater reliability; and (d) the maintenance and generalization of treatment results to participants' daily lives and the perceived value of those results (*social validity*) (Gersten et al., 2005; Horner et al., 2005).

### *Purpose of This Review*

In 2005, NJC members agreed that it was time to systematically review the past 20 years of communication intervention research involving persons with severe intellectual and developmental disabilities in light of today's standards of evidence-based practice. We felt that such a review of the evidence would afford useful information for providers who question the potential benefit of communication intervention for individuals with severe disabilities and would also identify directions for future research. Thus, we undertook this methodical review of the literature to address three broad and basic questions: (a) What are the characteristics of the research evidence that support the delivery of communication interventions to individuals with severe disabilities? (b) What is the nature and quality of the evidence? (c) How can these findings inform specific needs for future research?

Our review differs from other recent reviews of communication intervention research because of our broad focus on individuals with severe disabilities, the period of time addressed, and our reporting of findings that achieved acceptable interrater agreement by reviewers.

### **Method**

In this systematic literature review, we examined published communication intervention research conducted with individuals who had severe

intellectual and developmental disabilities during a 20-year period between 1987 and 2007. The initial search for the period of 1987 to 2006 was carried out by the National Center for Evidence-Based Research Practice in Communication Disorders of the American Speech-Language-Hearing Association. The search for the period of 2006–2007 was conducted by NJC members using the same procedure. We targeted this 20-year period because of the major improvements in special education laws that have affected individuals with severe disabilities and have served as a stimulus for applied research. Preparation for completing the review with acceptable reliability was a complex and time-consuming process; it was not possible to add research published after 2007. Only articles meeting the following criteria were selected for review: (a) were published in peer-reviewed journals, (b) were written in the English language, (c) had participants with severe disabilities, (d) were intervention studies dealing with language or literacy outcomes, (e) contained original data, and (f) were not case studies.

A four-step search process was applied to identify a pool of research articles meeting these requirements. First, 13 electronic databases were searched: CINAHL, Combined Health Information Database, ERIC, Education Abstracts, Exceptional Child Education Resources, Health Source: Nursing, Linguistics and Language Behaviour Abstracts, PsycARTICLES, PsycINFO, PubMed, Science Citation Index, ScienceDirect, and Social Science Citation Index. We used 31 search terms to select potential studies (e.g., augmentative or alternative communication [AAC], augmentative communication, communication, emergent communication); the full list is available on the NJC website. After this initial search, 47 expanded search terms were created and applied (e.g., Communication [MeSH Major Topic] and (augmentative OR alternative OR emergent OR nonsymbolic OR presymbolic OR intentional symbol\* OR speech generat\*). Third, the reference lists of all relevant articles identified were scanned for other possible studies. Finally, all publications authored by NJC members were searched. This search process generated a pool of 269 potentially relevant articles.

### *Development of the Research*

#### *Evaluation Instrument*

The NJC Evidence-Based Practices Data Entry Instrument (2008) was developed to evalu-

ate the characteristics of communication intervention research conducted with individuals who had severe disabilities. Using initial versions of the instrument during a face-to-face meeting, all 10 NJC members read and independently coded a subset of the 269 articles that were chosen to test the emerging inclusion criteria. Members discussed their ratings on the inclusion criteria, resolved their differences, and then revised the wording of items and replaced open-ended items with closed-ended items (i.e., yes/no and multiple choice). Second, we selected 50 articles at random from the pool and assigned 7 to 8 articles to each of 2 raters; these pairs of raters talked by telephone to compare their independent ratings on the inclusion criteria and the instrument items and to discuss and resolve differences. During regular conference calls, we made improvements in the coding form as a result of this work. Third, at a face-to-face meeting, the first 6 authors read 6 articles randomly selected from the pool of articles and independently rated the articles. The authors compared their ratings, discussed and resolved any differences, and then added information to the coding form that more precisely defined contested items. For example, we added developmental ranges for judging whether participants had severe disabilities and for identifying their pretreatment communication characteristics (e.g., "follows simple verbal or gestural directions" was supplemented with "and/or reported receptive language age of 9 to 18 months"). At the end of this meeting, we finalized the instrument.

The instrument was converted into an electronic coding form and placed on a web-based survey platform that can be found online (<http://www.Zoomerang.com>). This format allowed easier data entry by reviewers in different locations. The final version of the instrument consisted of 39 questions, 32 of which addressed the content of the studies meeting inclusion criteria (available on the NJC website <http://professional.asha.org/njc>). The instrument was divided into four sections: (a) reviewer/article information (2 items), (b) inclusion criteria (5 items), (c) study description and characteristics (29 items), and (d) summary of evidence quality (3 items). The last 3 instrument items on quality of evidence used a rating system developed by the National Research Council (2001) for evaluating the quality of evidence provided by any one study on the basis of three key indicators: internal validity, external validity, and generalization. Eighteen of the 32 content

questions (56.2%) required reviewers to make a single choice. The remaining 14 items (43.8%) required that the reviewer check all of the 4 to 8 options that applied to a given study. For example, under Item 12, diagnoses/disabilities of participants with severe disabilities, the reviewer checked any of the eight options that characterized the study participants who had severe disabilities (e.g., developmental delay or intellectual disability, cerebral palsy, autism spectrum disorders, sensory impairments). Thus, although the instrument had 32 numbered content items, each study was coded on a total of 104 items, including all the multiple option subitems. For these items where multiple options could be checked, frequency percentages could exceed 100%.

### Review Procedure

The pool of 269 potentially relevant articles identified through the search was divided evenly among the first 6 authors who read, coded, and entered their ratings for every assigned article into the instrument on the web-based survey platform. They examined each article in this pool using a stepwise process; articles judged to not meet all three inclusion criteria were not evaluated beyond these inclusion items, and articles judged to meet all three inclusion criteria were included in the database and evaluated using the entire instrument. To be included in the database, each article had to meet three inclusion criteria and, thus, be judged as a study that (a) described an intervention, (b) included one or more participants of any age with severe disabilities, and (c) applied an intervention addressing one or more areas of communication performance. The definition of *severe disabilities* that was applied in the second criterion allowed for multiple components and included a broad description, along with a specific IQ cutoff (i.e., 44 or below, allowing for measurement error), and language age guidelines that were aligned with chronological age (CA) (Item 5 of the NJC instrument). For the third criterion, *communication performance* was defined as learning to understand and/or produce communication messages to a communication partner using any mode, including graphic, natural gestures, sign language, speech, picture symbols, and addressing one of the following functions: requesting, commenting, protesting, conveying social niceties, answering questions, repairing after a breakdown (Item 6). Thus, we did not review

articles that focused on the component skills of communication, such as matching-to-sample or picture identification, unless the study included an aspect of teaching the participant(s) to use the component skills to communicate with another person.

From the pool of 269 potentially relevant articles identified through our search, we judged 116 studies (43%) to meet the three inclusion criteria and, thereby, qualify for further review on the 32 content items on the instrument. The findings reported in this review were drawn from this qualified database of 116 studies.

### *Interrater Agreement*

Interrater agreement was assessed in two ways to calibrate multiple raters with each other and to assure that raters were accurate and procedures of the review were replicable (Bakeman & Gottman, 1997). The first type of interrater agreement addressed concurrence in judging the fulfillment of the inclusion criteria of an article. The second type addressed concurrence in rating the remaining 32 items on the instrument.

*Inclusion criteria items.* To calculate interrater agreement for scoring studies on the three inclusion criteria, we selected a group of 71 articles (26.4%) at random from the pool of 269 potentially relevant articles. Each article was randomly assigned to 2 reviewers (primary and secondary) for independent rating. The primary reviewer's ratings on the inclusion criteria items for each article were compared with the secondary reviewer's ratings on a point-by-point basis and scored for exact agreement or disagreement. An agreement percentage was calculated by dividing the total number of agreements by the number of agreements plus disagreements and multiplying this figure by 100. After comparison, the primary reviewer's ratings were retained and the secondary reviewer's ratings were dropped. The interrater reliabilities for the inclusion criteria were as follows: (a) investigator(s) describes an intervention or treatment (95.8%), (b) investigator(s) includes one or more participants with severe disabilities (84.5%), and (c) treatment addresses one or more areas of communication performance (81.7%).

Differences in scoring for the second criterion appear to have been due to researchers omitting or indistinctly reporting disability information, such as IQs, IQ range labels (e.g., moderate intellectual disability, severe intellectual disability), or pretreat-

ment communication assessment results. Differences in scoring for the third criterion seem to have been caused by inadequate information provided on the researcher(s)' focus; the required focus was to teach a person to understand and/or produce communication messages to a communication partner using any mode and addressing one or more basic functions. In some of the research with disagreement on this item, the investigator(s) addressed component skills, such as matching symbols or letters to objects. In some cases it was difficult to determine whether the intervention included using these component skills to communicate to another person.

*Content items.* Second, we assessed interrater agreement on the 32 items on the instrument and the associated subitems describing the study. Fourteen items had a multiple choice format and required reviewers to check all options that applied; the remaining 18 items had a yes/no format. Thirty-five studies (30.2% of the qualified database) were randomly selected from the 116 studies that met inclusion criteria. These studies were randomly assigned to 2 reviewers (primary and secondary) for independent rating. As with the inclusion criteria, the primary reviewer's ratings on the content items for each study were compared to the secondary reviewer's ratings on a point-by-point basis and scored for exact agreement or disagreement. An agreement percentage was calculated by dividing the total number of agreements by the number of agreements plus disagreements and multiplying this figure by 100. After comparison, the primary reviewer's ratings were retained and the secondary reviewer's ratings were dropped.

Moderate to strong interrater agreement was attained on all three inclusion criteria and on 26 of 32 content items (81.3%) or 92 of 104 total content items (88.5%), including instrument subitems. The 12 out of 104 instrument subitems that fell below levels of 70% were dropped from further analyses (i.e., Items 17.5, 18.2, 18.3, 21.4, 21.7, 30.3, 30.6, 32, 33, and 37-39 on the NJC instrument). The results presented below include only those for items on which our independent raters achieved agreement of at least 70%.

## **Results**

### *Characteristics of the Research*

*Number, gender, and CAs of participants.* We identified number, gender, and CAs of the

participants with severe disabilities in each study and reported on 461 individuals (287 males, 174 females) with severe intellectual and developmental disabilities in the 116 studies in the database, with a range of 1 to 41 participants and a mean of 4.0 participants. Their average age was 13.7 years.

When a study included one or more participants judged to have severe disabilities and other participants who did not, we included only the information for participants with severe disabilities. Table 1 shows the distribution of participants by age group and other participant characteristics.

**Table 1.** Characteristics of Participants With Severe Intellectual and Developmental Disabilities

Characteristic	Studies reporting participants of each type <sup>a</sup>	
	<i>n</i>	%
CA level (in years)		
0–5	51	44.0
6–11	42	36.2
12–17	33	28.4
18–20	23	19.8
21 and older	29	25.0
Participant disability		
Intellectual disability	92	79.3
Autism	53	45.7
Multiple disabilities	40	34.5
Cerebral palsy	21	18.1
Specific syndrome <sup>b</sup>	19	16.4
Sensory impairment	16	13.8
Behavioral disorder	9	7.8
Pretreatment communication level		
Prelinguistic	77	66.4
Emerging	60	51.7
Multiple, nonecholalic words	7	6.0
Other (e.g., echolalia)	11	9.5
Pretreatment communication mode		
Speech	57	49.1
AAC <sup>c</sup>		
Unaided	20	17.2
Aided	10	8.6
Aided, w/speech output	5	4.3
Gestures/vocalizations	69	59.5
Other <sup>d</sup>	25	21.6
Pretreatment receptive communication ability		
Not responsive: RLA <sup>e</sup> ≤ 9 months	13	11.2
Simple directions: RLA 9–18 months	37	31.9
Single words: RLA 18–30 months	26	22.4
Grammar/syntax: RLA > 30 months	6	5.2
Not reported	62	53.4

<sup>a</sup>Out of 116 studies included in this review, data were reported only for study participants who could be identified with severe intellectual and developmental disabilities; numbers total greater than 116 because some studies included participants of different ages, diagnoses, and/or communication abilities. <sup>b</sup>Specific syndromes reported: Down (17), Retts (5), and one each for Klinefelter's, microcephaly, Pierre Robin, fragile X, and Angelman. <sup>c</sup>Augmentative and alternative communication. <sup>d</sup>For example, behavior problems. <sup>e</sup>Receptive language age.

As shown in the table, a larger number of the 116 studies reported interventions with younger children (80% included participants younger than 12 years) than with adults (only 25% included one or more participants 21 years or older).

*Disability and communication characteristics of research participants.* We identified the primary diagnoses or disabilities of participants in each study, noting any specific genetic disorders or syndromes mentioned. Table 1 summarizes the number of studies including one or more participants within each of the coded disability categories. Nearly 80% of the studies included at least one participant with a diagnosis of intellectual disability, whereas only 19 studies (16.4%) reported participants with specific genetic disorders or syndromes. Down syndrome (17 studies) and Rett syndrome (5 studies) were the most frequently identified syndromes. Only 9 of the selected studies included one or more participants identified as having a behavioral disorder. Forty of the 116 studies (34.5%) included participants who had the label of multiple disabilities. In 66.4% of the studies, researchers identified two or more disability categories as being reflected in their sample, with the highest number being six disabilities.

*Pretreatment communication levels of participants.* We coded the reported communication levels and modes of participants prior to intervention. For most of the participants in these studies, as shown in Table 1, pretreatment expressive communication was described as being prelinguistic (66.4%; no real words in any mode or reported expressive language age of less than 18 months) or emergent (51.7%; reported expressive language age between 18 and 30 months). Only a small number of participants communicated at a multiword, non-echolalic level. The most common modes of communication reported were gestures and speech. For participants who were reported to use some type of AAC prior to intervention, the most common type was unaided AAC, typically manual signs.

In terms of the pretreatment receptive language or comprehension abilities of participants in these studies, we found that half of the study authors provided no information at all about this aspect of participants' communication abilities (Table 1). When receptive communication skills were reported, the most common levels described were "follows simple directions"; "receptive lan-

guage age (RLA) of 9–18 months"; or "understands a few single words; RLA 18–30 months."

*Dependent variables.* Researchers measured several aspects of communication performance that were explicitly targeted as outcomes of treatment. Table 2 summarizes the types of communication performance that were measured as outcomes of treatment in these 116 studies. By far the most frequently targeted outcome was improvement in expressive communication followed by improvement in interaction or conversation. The most frequently targeted expressive communication mode was speech, followed closely by AAC device with no speech output, and then AAC device with speech output. More than one mode was assessed in 43.5% of the studies, with a range of 1 to 4 modes. Again, we found that in most of these studies, researchers did not target or measure receptive communication in any mode as a targeted outcome of intervention. Of those studies that did include a receptive communication dependent variable, the most common mode measured was understanding a partner's spoken speech. The most frequent outcome measure of communication function reported in these studies was "regulate the behavior of others (e.g., requesting, rejecting)." We found that researchers in 18.1% of these studies identified more than one communication function as the targeted outcomes of the experimental intervention.

In addition to assessing communication performance, in 10.8% of the studies, researchers measured challenging behavior during baseline and intervention and reported their findings on these behaviors in their results.

*Independent variables.* We coded the following characteristics of the specific intervention(s) applied in the study: (a) location(s) where intervention was conducted, (b) the instructional methods used (e.g., individual, group, distributed trial, decontextualized), and (c) the person(s) who delivered the intervention. Table 3 displays the distribution of these procedural features among the 116 studies. The most commonly used setting in these studies was the classroom, followed by pull-out environments (e.g., therapy rooms or experimental rooms), home, and community. In about a third of the studies, researchers reported other settings (e.g., playground, empty classroom, conference room, cafeteria, group home) or the setting was not clearly specified. In many studies

**Table 2.** Dependent Variables Measured as Intervention Outcomes

Data input instrument item	Studies reviewed <sup>a</sup>	
	<i>n</i>	%
Aspects of communication performance measured as outcome		
Expressive communication/expressive language	94	81
Comprehension/receptive language	8	6.9
Interaction/conversation	27	23.3
Other	11	9.5
Expressive communication mode target		
Speech	48	41.4
ACC <sup>b</sup>		
Unaided	25	21.6
Speech output	29	25
Aided, no speech output	42	36.2
Nonsymbolic (gestures, vocalizations)	NR <sup>c</sup>	NR
NA/not measured	8	6.9
Receptive communication mode target		
Speech	11	9.5
ACC		
Unaided	3	2.6
Speech output	6	5.2
Aided, no speech output	6	5.2
Gestural or contextual	4	3.4
NA/not measured	96	82.8
Communication function target		
Regulate behavior	62	53.4
Engage another	25	21.6
Establish joint attention	19	16.4
NA/not measured	37	31.9
Interaction/conversation target		
Turn-taking	13	11.2
Joint attention	11	9.5
Imitation	8	6.9
Initiation/spontaneous	NR	NR
Repair	1	1
Topic maintenance	2	1.7
NA/not measured	NR	NR
Other	16	13.8

<sup>a</sup>Out of 116 studies included in the review, data reported are only for study participants who could be identified with severe intellectual and developmental disabilities; numbers total greater than 116 because some studies included participants of different ages, diagnoses, and/or communication abilities. <sup>b</sup>Augmentative or alternative communication.

<sup>c</sup>Data not reported because interrater reliability among article reviewers was < .70.

(32.8%), researchers delivered intervention in more than one setting.

In most of the studies, the intervention was delivered to participants on an individual or one-on-one basis, with group interventions occurring

in only about 10% of the research (see Table 3). Teaching trials were distributed over an activity or session, rather than massed into a short time segment, in nearly half of these studies. In 39.6% of studies, researchers delivered the experimental



**Table 3.** Characteristics of Study Interventions

Data input instrument item	Studies reviewed <sup>a</sup>	
	<i>n</i>	%
<b>Intervention setting</b>		
Classroom	51	44
Home	32	27.6
Pull-out	40	34.5
Community	6	5.2
Other	34	29.9
<b>Implementation method</b>		
Individual/one-to-one	102	87.9
Group	11	9.5
Massed trial	NR <sup>b</sup>	NR
Distributed trial	53	45.7
Decontextualized	46	39.6
Contextualized	NR	NR
Other	4	3.4
<b>Person(s) delivering intervention</b>		
Parent	19	16.4
Teacher	41	35.3
Paraprofessional	14	12.1
Speech language pathologist	7	6
Peer	11	9.5
Experimenter	60	51.7
Other	23	19.8

<sup>a</sup>Out of 116 studies included in the review, data reported are only for study participants who could be identified with severe intellectual and developmental disabilities; numbers total greater than 116 because some studies included participants of different ages, diagnoses, and/or communication abilities. <sup>b</sup>Data not reported because interrater reliability among article reviewers was < .70.

intervention in decontextualized settings that were removed from the natural communication environment, with conditions manipulated according to time, setting, or individuals present. We defined *decontextualized settings* as created treatment conditions that were strikingly different from scheduled routines. In the majority of studies, the intervention was delivered by an experimenter, and by others in a decreasing order of frequency: classroom teacher, parent, paraprofessionals, peers, or speech-language pathologists. In the category "other" individuals who delivered the intervention were, for example, classroom

staff member, graduate student, direct services staff member, coworker, or occupational therapist, but half of the "other" group was not clearly specified. More than one individual delivered intervention to participants in 35.3% of the studies.

### *Nature of the Evidence*

*Primary outcomes of the intervention.* In 95.7% of the 116 studies reviewed, researchers reported immediate, positive results in the target skill following intervention. This item was judged by examining reported changes against time, depending upon the experimental design (e.g., graphs were examined for single subject research). Although we did not specify criteria for immediate positive results, our interrater agreement on this item was 88.89%.

We did not categorize research on the specific intervention methods applied. However, as noted previously, the majority of these interventions focused on either improving expressive language or interaction skills (see Table 2). To achieve this focus, a wide range of interventions were reported (e.g., Picture Exchange Communication System, functional communication training, systematic social interactive training, teaching conversational exchanges with peer partners and communication books, Enhanced Milieu Teaching, using visual supports to teach initiations, application of object and movement cues to teach receptive skills, reinforcement strategies to teach signing, time delay to promote speech). Because we did not identify or classify intervention methods, it is not possible to describe their frequency of use, compare their effects, or analyze which interventions or combinations of interventions were associated with stronger outcomes for participants. What was found from this review is that the majority of these 116 intervention studies (95.7%) were reliably judged as achieving immediate positive results or measurable improvement in one or more aspects of communication performance in participants with severe disabilities.

Finally, we coded each study for research design and validity (e.g., inter- and intrarater reliability, treatment fidelity, social validity) and on characteristics of intervention effectiveness (e.g., immediate results, long-term effects). Regarding experimental design, of the 116 studies reviewed, experimental single subject research designs were used in 67.2% of the studies, while

quasi-experimental designs were used in 19%, qualitative designs were used in 9.5%, and experimental group designs were used in only 3.4% of the studies.

**Generalization.** We examined whether the 116 studies measured any sort of generalization, including *stimulus generalization* (i.e., the transfer of target skills to, for example, new partners, materials, environments) or *response generalization* (i.e., changes in behaviors similar to those targeted). In a little more than half of the studies (51.3%), researchers included some measure of skill generalization (e.g., to new partners, settings). Consistent with these findings, several conditions were reported in this database that may have contributed to the promotion of stimulus generalization: (a) more than one individual delivered intervention to participants in 35.3% of the studies; (b) the most commonly used setting for intervention in these studies was the classroom (44%) rather than an artificial setting, and (c) in many studies (32.8%) intervention was delivered in more than one setting.

**Maintenance.** Less information was reported in these studies about skill maintenance or the continued performance of target behavior after intervention was withdrawn. In only 29 of these 116 studies (25.2%) did investigators report measuring maintenance of effects 3 or more months after all intervention was completed. In the vast majority of the communication intervention research we reviewed (74.8%), researchers did not measure the maintenance of the target skills 3 months or longer following intervention.

### *Quality of the Evidence*

Interrater agreement was measured in almost all of the 116 studies (89.5%); however, researchers measured intrarater agreement (i.e., evidence that raters were consistent over time, included test-retest reliability within a single individual) in only 2.6% of the studies. In 32.2% of the studies reviewed, investigators assessed *fidelity of treatment* (i.e., evidence that experimental conditions were implemented as described). Finally, researchers assessed some feature of *social validity* (i.e., any measure of acceptability or benefit of the intervention from the perspective of experts or individuals who interact with the participant) in 16.8% of the research.

Some researchers who measured interrater agreement also measured various combinations of

these research characteristics: treatment fidelity, social validity, generalization, or maintenance. At least two of these four research characteristics were measured in 32.8% of the studies, whereas three of these four characteristics were measured in only 7.8% of the studies. All four characteristics were measured in only 2.6% of the studies. The measurement of generalization and treatment fidelity was the most frequent combination (20.7% of the studies assessed both).

## Discussion

Our three purposes in this review were to (a) identify the characteristics of the research evidence that supports the delivery of communication interventions to individuals with severe disabilities, (b) describe the nature and quality of that evidence, and (c) suggest how these findings inform future research. The evidence reviewed indicates that positive changes in some aspects of communication were reported in nearly all of the studies in the database.

### *Characteristics of the Research*

In this literature review we identified 116 research studies published between 1987 and 2007 that described an intervention addressing the communication performance of at least one individual with severe intellectual and developmental disabilities. The typical study applied single subject experimental design (67.2%) with a mean of 4 school-aged participants (mean age 13.7 years) with intellectual disability (79.3%). The participants' typical pretreatment expressive level was reported as being prelinguistic (66.4%) or emerging (51.7%), whereas nonsymbolic gestures and vocalizations were their most frequent communication mode (59.5%). Intervention characteristically was delivered in the classroom or in pull-out settings. In most studies intervention was delivered on a one-to-one basis (87.9%), often using distributed trials (47.5%). Participants' improvement in expressive communication was the most frequently measured outcome (81%), and researchers reported immediate positive results in the target skills following intervention in 95.7% of the studies reviewed.

Speech and various forms of AAC were the most frequently targeted communication modes. Consistent with current recommendations to provide multi-modal communication,

in 43.5% of the studies researchers targeted and measured more than one mode (Beukelman & Mirenda, 2005). Of these AAC modes, communication using devices with no speech output (e.g., picture communication books, picture symbols as in Picture Exchange Communication System) was targeted and measured most often (32.2%), whereas communication with speech-generating devices (e.g., Wolf communication board, Introtalker) or unaided AAC (e.g., signing) were addressed about the same amount but less often than AAC modes with devices and no speech output. Given the pretreatment characteristics of the participants, it was not surprising that when communication function was measured, more than half of the researchers assessed regulating the behavior of others, as in requesting. What seemed surprising, however, was that in 33% of the studies, researchers did not report measuring any communication function.

Our first purpose in this literature review was to determine what research evidence there is that supports and describes the delivery of communication interventions to individuals with severe disabilities. In this review, we identified 116 studies in which researchers specifically addressed this question using some type of experimental or quasi-experimental design. In almost all of these studies (95.7%), investigators reported that the intervention was followed by positive and immediate results for most or all participants with severe disabilities. These overwhelmingly positive outcomes are partly due to selection biases in publications; that is, only studies with positive outcomes tend to be submitted and accepted for publication (Torgerson, 2006). However, the published evidence clearly supports the provision of intervention services to improve the communication skills of children and adults with severe disabilities.

### *Nature and Quality of the Research*

Our second purpose in this review was to describe the quality of the evidence base or the believability of the research findings. Specifically, we were interested in learning the degree to which these 116 studies incorporated design features that assure the *internal validity* of findings (i.e., whether the results are attributable to the experimental procedures as described) and the *external validity* of findings (i.e., whether the results are useful and

generalizable to other members of the target population). The review process included consideration of recent guidelines for evaluating research (e.g., Gersten et al., 2005; Horner et al., 2005; Justice & Snell, 2007; Lonigan, Elber, & Johnson, 1998; National Research Council, 2001; Odom et al., 2005). Although these documents were beneficial, there were a number of challenges in applying many of the so-called "gold standards" to the studies we reviewed.

*Summative rating of research quality.* The first challenge was attempting to give a summative rating for the quality of each research study. Rather than tease apart all elements that make up internal and external validity (e.g., Troia, 1999; Tuckman, 1999) and judge each study on all of these elements, we chose instead to include three instrument items (internal validity, external validity, and generalization) from the National Research Council's (2001) frequently cited review of the evidence on education of children with autism. It is worth noting that we were not able to achieve acceptable reliability on these items. In the original National Research Council's (2001) report, different contributors were assigned to conduct reviews of research in various intervention areas (e.g., sensory, motor, social). In order to produce comparable data from these diverse literatures, these reviewers were all instructed to use the same scale to rate each study in their topic area. During our reliability training, we discussed these items at some length in an attempt to resolve initial disagreements; but we did not modify the actual wording of the items because we wanted to be able to compare our findings to those in the National Research Council report. When we completed our review and realized that we had not achieved acceptable interrater reliability on these three scales, we reviewed the report to learn how they applied these scales, only to find no mention of rating accuracy or reliability.

We had the opportunity to ask one of the original National Research Council (2001) review authors whether there was any formal reviewer training or instruction to assure consistent application of the scales; she confirmed that there was not (G. T. Baranek, personal communication, January 27, 2010). This is not surprising, in hindsight, because the National Research Council report was completed almost 10 years ago, when our current sensitivity to the issue of interrater reliability (or even intrarater reliability) in meta-

analyses or summative reviews of research literature was less fully developed.

As reported earlier, however, we did reliably code each study on some specific elements of internal validity (experimental design, treatment fidelity, operationalized measures, interrater agreement) and external validity (participant characteristics and background, disability, and generalization). Although interrater agreement was reported in 89.5% of the studies and dependent measures were identified and described, one serious limitation of internal validity concerned treatment fidelity; only 32.2% of the researchers assessed whether experimental conditions were being implemented as described. Of the specific external validity issues that we assessed, we found that reliable documentation of generalization was absent in half of the research, although participant characteristics and disability were reported.

*Experimental design.* A second challenge in applying “gold standards” to the studies we reviewed concerned experimental design. The standards for judging quality group and quasi-experimental research (Gersten et al., 2005) differ from those for judging quality single subject research (Horner et al., 2005; Kennedy, 2005), and all three types of design were found in the database. We reliably identified the type of experimental design used in each study and the reported intervention results; however, based on the fact that all studies were published in peer-reviewed journals, we did not evaluate how well each study met specific design standards. In two thirds of the studies (67.8%) meeting inclusion criteria, researchers used single subject research designs and involved a small number of participants in a given study. The choice of single subject design suits the low-incidence and heterogeneous nature of the population of individuals with severe disabilities (Horner et al., 2005). By using the unit of the individual for analysis as well as for delivery of the intervention, single subject design enables identification of causal or functional relationships “without requiring the assumptions needed for parametric analysis (i.e., normal distribution)” (p. 173). Given the abundance of single subject research that we identified, future reviewers should categorize communication interventions and then conduct meta-analyses of the single subject research so as to identify the credibility of specific intervention procedures.

In this review we identified only 4 studies (3.5%) meeting the inclusion criteria in which a treatment group was compared with a control or contrast group. Although inadequate information was given to calculate an effect size for one of these studies, we calculated the effect size for the remaining three (i.e., Girolametto, 1988; Girolametto, Weitzman, & Clements-Baartman, 1998; Yoder & Layton, 1988). Girolametto et al. used a Mann-Whitney  $U$  statistic, in which the difference in median scores can indicate the effect size. The difference in medians in their study was 4.5 words, indicating that parents reported, on average, 4.5 more words learned by children in the treatment group than by those in the control group. Girolametto (1988) studied the effects of training mother-child dyads in the use of a social conversational approach to a control group. Children of mothers in the experimental group had a higher turn-taking ratio, took more verbal turns, and exhibited a more diverse vocabulary than did the control group children, yielding large effect sizes for the experimental group on all three variables (Cohen's  $d$  of .85, .84, and .84, respectively). Finally, Yoder and Layton (1988) compared child-initiated speech in children with autism under four different treatment conditions—speech alone, sign alone, simultaneous speech and sign, and alternating speech and sign. Less speech was produced in the sign only condition than in any of the other three conditions (large effect,  $d = .707$ , when compared to speech alone; medium effect,  $d = .585$ , when compared to simultaneous speech and sign; and small effect,  $d = .33$ , when compared to alternating speech and sign). In summary, 3 of the 4 total group design studies in this database (treatment group compared with control or comparison intervention) demonstrated moderate to strong effect sizes in their application of a communication intervention to individuals with severe disabilities. Although group comparison research was only 3.5% of the database, these positive effects are consistent with the overall supportive nature of the evidence.

*Reliable ratings.* A third challenge came from the task of reliably judging research characteristics across a disparate literature base using multiple reviewers. Despite these conditions, we reliably evaluated 26 of 32 content items (81.3%) or 92 of 104 total content items (88.5%), including instrument subitems. These items addressed many aspects of the state of the current evidence in interventions about communication for individu-

als with severe disabilities. The 12 out of 104 instrument subitems that fell below 70% and were dropped from further analyses included identification of several implementation methods (massed trial, contextualized intervention), frequency and duration of training, and the three National Research Council summary items on research quality. Reliable items far outweighed those not achieving reliability. It is important also to note the frequent omission of reliability reporting in literature reviews in the field of education and psychology, either for inclusion criteria or for characteristics of the reviewed research.

### *Specific Needs for Future Research*

Our final purpose in this literature review was to determine how these findings can inform specific needs for future research.

*Treatment fidelity.* One critical aspect of internal validity is a measure of the fidelity of treatment implementation. Documentation of fidelity of a study to its treatment protocol provides an essential measure of the consistency with which the independent variable(s) used in the experimental intervention was actually applied (Tuckman, 1999). Because the independent variable is applied over time in single subject research, repeated measurement of the fidelity of implementation is the accepted approach for documenting adequate consistency of implementation (Gresham, Gansle, & Noell, 1993; Horner et al., 2005). Furthermore, because many communication interventions consist of multiple strategies (e.g., timed prompting, precise error correction, contingent praise, environmental arrangement) delivered with a prescribed frequency, assessing the fidelity of implementation provides a measure of confidence in the independent variable(s) and contributes to the determination of what procedures are accountable for treatment effects.

Only about 30% of the 116 studies we reviewed contained reports of any measure of treatment fidelity; this finding is similar to those reported in other communication review papers (Gresham et al., 1993; Howlin, Magiati, & Charman, 2009; Hwang & Hughes, 2000; Schlosser & Lee, 2000; Snell, Chen, & Hoover, 2006). Documentation of fidelity of implementation is an essential requirement of quality research (e.g., Gersten et al., 2005; Horner et al., 2005). Including treatment fidelity measures in future studies is of paramount importance.

*Generalization and maintenance.* In this review, we found that researchers assessed generalization only half of the time, and maintenance of effects and social validity were measured less than one fourth and one sixth of the time, respectively. It is of value to practitioners to know whether an intervention can produce communication skills that will transfer beyond instructors, instructional setting, or the specific forms taught and that will endure over time. Interventions that have longer term effects on participants' communication are potentially more valued by teachers, speech-language pathologists, and parents than are interventions with short-lived effects. Although researchers in one third of the studies employed more than one interventionist and more than one instructional setting, we found that the measurement of skill generalization (reported half the time) was still deficient in this database. We also found that when long-term effects (maintenance at least 3 months following intervention) were measured, they were reported as being successful most of the time; however, long-term effects were reported in only about one fourth of the research reviewed.

An examination of other recent communication reviews results in similar deficiencies. Schlosser and Lee (2000) assessed the specific types of generalization assessments in a group of 50 AAC intervention studies and found that generalization across persons and across settings was reported about one third of the time, while generalization across stimuli was reported half as often. Hwang and Hughes (2000) found that 9 of the 16 prelinguistic intervention studies (56%) they reviewed included some measure of generalization, whereas only 6 of the 16 studies (38%) reported follow-up or maintenance data. Snell et al. (2006) found that 40% of the AAC studies they reviewed assessed generalization, but only 5% assessed and found maintenance of effects.

*Identification of distinct intervention practices.* In this review we identified intervention settings, interventionists, and some characteristics of implementation (e.g., one-to-one, group). However, we did not attempt to classify interventions by their specific treatment components as some reviewers have done (e.g., Hepting & Goldstein, 1996; Snell et al., 2006). Hepting and Goldstein found that terms used by some researchers to describe naturalistic language interventions often were inconsistent. Goldstein (2002) elaborated: "The under-specification of what instructional

procedures are active in interventions represents a large obstacle to those interested in conducting treatment comparisons" (p. 391). Not only do treatment components need improved specification, but those using the research findings must not assume that treatment components called by the same name are equal. We agree with Hepting and Goldstein's recommendation that communication researchers specify their treatment components in more detail, perhaps using a taxonomy of treatment components.

*Treatment intensity and duration.* We tried to identify for each study the intensity or dosage of intervention (i.e., rating options started with 2 or more times daily, 7 days/week, and ended with less than once a week) and the duration of intervention (i.e., rating options started with less than 1 month and ended with more than 2 years). Treatment intensity and duration proved very difficult to determine in these 116 studies, in part, because there is no standard way of reporting this information. Regarding duration, intervention data were sometimes reported in terms of trials or sessions to criterion or total number of trials, yet the number of trials or sessions completed per day or week was not specified. Thus, an interventionist wishing to replicate reported results would not be able to predict the amount of therapy, timing of intervention (e.g., number of sessions/opportunities per day), or the length of time anticipated to achieve reported outcomes.

In a recent review of early intensive behavioral interventions for children with autism, Howlin et al. (2009) revealed similar problems, with few researchers reporting the actual hours of intervention or providing clear information on the length of time children were involved in the intervention. Describing the intensity and duration of intervention is crucial in characterizing an intervention and its effects (Goldstein, 2002). School systems and parents are less likely to select methods that, though shown to be effective, require extensive instructional time (Goldstein, 2002; Mirenda, 2001). Over 30 years ago, Connell, Spradlin, and McReynolds (1977) recommended that clinicians refuse to use communication programs unless adequate information was provided to support the usefulness of a program (i.e., specific descriptions of individuals on whom the program was tested, trials to criterion for each program step, percentage of students completing each program step, evidence of generalization). Although accurate reporting of treatment intensi-

ty is simpler than evaluating the effects of delivering treatment with different intensities or over varying lengths of time, researchers must first report their treatment intensity and duration before treatment efficiency can be evaluated.

*Description of research participants.* We found it challenging to define the inclusion criterion for studies that "include one or more individuals, of any age, with severe disabilities" (National Joint Committee, 2008, p. 1). Definitions of *severe disabilities* typically do not contain quantifiable characteristics, but cite the extremely heterogeneous nature of the group. Thus, early in instrument development, we began with a broad definition and the principle that we would "err on the side of inclusion." Our agreement on this inclusion criterion improved when we added other options to the broad definition: a specific IQ cutoff and language age guidelines that were aligned with CA and could be used if IQs were not provided in a study.

The predominant pretreatment characteristics reported for the 467 participants with severe disabilities in the database were consistent with performance levels reported by others for individuals with severe disabilities prior to intervention (Beukelman & Mirenda, 2005; Paul & Wilson, 2009). Participants' expressive communication levels rarely included word combinations and typically were emergent or prelinguistic; individuals were described most often as using gestures and nonsymbolic modes, speech, and problem behavior to communicate, whereas AAC modes were infrequent before intervention. Participant descriptions varied widely, with some researchers relying on narrative depiction and others reporting primarily standardized assessment information. This lack of complete and comparable participant descriptions has been noted in several reviews of communication interventions for children with autism (Goldstein, 2002; National Research Council, 2001) and in a recent position paper, in which Tager-Flusberg et al. (2009) advocated that a developmental framework be used to define spoken language benchmarks for children with autism. Specifically, we support Goldstein's recommendation that "the field would benefit from a set of conventions that would help standardize the sharing of descriptive information about participants" (p. 390).

*Limitations.* Finally, this review has several limitations. First, as discussed, we were unable to achieve interrater agreement on all items in the

instrument, including the three summative ratings of research quality taken from the National Research Council (2001) report. Second, we did not identify the distinct intervention components used by the researchers in each study. Third, we were unable to include research from the most recent year, 2008.

## Conclusions

The most compelling finding in this systematic review was its clear support for the success that individuals with severe disabilities can have in learning a broad range of expressive or interactive communication when they are provided with systematic intervention.

To advance toward evidence-based practices in communication intervention for individuals with severe disabilities, researchers must carry out a higher quality of research than generally has been evident over the past 20 years. This means that researchers must first define their participants in more thorough and standard ways. Furthermore, they need to document acceptable fidelity of implementation. Their tests of the intervention should include an assessment of generalization to another setting and measurement of maintenance beyond the end of experimental intervention. Finally, researchers need to describe their interventions methodically, including setting, interventionist, methods of implementation, treatment intensity and duration, as well as identification of the specific components of the intervention. With these improvements, it will be possible to assess the evidence base of practices that yield predictable positive effects on the communication of individuals with severe disabilities.

## References

- American Physical Therapy Association. (n.d.). *Evidence-based practice* (first paragraph). Retrieved from <http://www.asha.org/docs/html/PS2005-00221.html>
- American Psychological Association. (2005). *Policy statement on evidence-based practice in psychology*. Retrieved from <http://www.apa.org/practice/ebstatement.pdf>
- American Speech-Language-Hearing Association. (2005). *Evidence-based practice in communication disorders* [Position Statement]. Retrieved from <http://www.asha.org/docs/html/PS2005-00221.html>
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.
- Beukelman, D. R., & Mirenda, P. (2005). *Augmentative and alternative communication* (2nd ed.). Baltimore: Brookes.
- Committee on Educational Interventions for Children With Autism. (2001). *Educating children with autism*. Washington, DC: National Academy Press.
- Connell, P. J., Spradlin, J. E., & McReynolds, L. V. (1977). Some suggested criteria for evaluation of language programs. *Journal of Speech and Hearing Disorders*, 42, 563-567.
- 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, 149-164.
- Girolametto, L. E. (1988). Improving the social-conversational skills of developmentally delayed children: An intervention study. *Journal of Speech and Hearing Disorders*, 53, 156-167.
- Girolametto, L. E., Weitzman, E., & Clements-Baartman, J. (1998). Vocabulary intervention for children with Down syndrome: Parent training using focused stimulation. *Infant Toddler Intervention: The Transdisciplinary Journal*, 8, 109-125.
- Goldstein, H. (2002). Communication intervention for children with autism: A review of treatment efficacy. *Journal of Autism and Developmental Disorders*, 32, 373-396.
- Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, 1, 297-306.
- Guttman, S. A. (2009). *How to appraise research: Elements of sound applied design*. (Editorial). *American Journal of Occupational Therapy*, 63, 123-125.
- Hepting, N. H., & Goldstein, H. (1996). What's natural about naturalistic language intervention? *Journal of Early Intervention*, 20, 249-265.
- 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, 165-180.
- Howlin, P., Magiati, M., & Charman, T. (2009). Systematic review of early intensive behavior-

- al interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities*, 114, 23–41.
- Hwang, B., & Hughes, C. (2000). The effects of social interactive training on early social communicative skills of children with autism. *Journal of Autism and Developmental Disabilities*, 30, 331–343.
- Justice, L. M., & Snell, M. E. (2007). Illustration of a three-step process for identifying the level and quality of empirical support for manualized treatments. *Evidence-Based Communication Assessment and Intervention*, 1, 37–47.
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.
- Lonigan, C., Elber, J., & Johnson, S. (1998). Empirically supported interventions for children: An overview. *Journal of Clinical Child Psychology*, 27, 138–145.
- Mirenda, P. (2001). Autism, augmentative communication, and assistive technology: What do we really know? *Focus on Autism and Other Developmental Disabilities*, 16, 141–151.
- National Joint Committee for the Communication Needs of Individuals With Severe Disabilities. (2008). *NJC Evidence-Based Practices Data Entry Instrument*. Unpublished document. Available from <http://www.asha.org/docs/html/PS2005-00221.html>
- National Research Council. (2001). *Educating children with autism* (Committee on Educational Interventions for Children With Autism, Division of Behavioral and Social Sciences and Education). Washington, DC: National Academy Press.
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practice. *Exceptional Children*, 17, 137–148.
- Office of Special Education Programs. (OSEP). Technical Assistance Development System (TADS). (1985). *A symposium: Children and youth with severe handicaps—Effective communication*. Washington, DC: Author.
- Paul, R., & Wilson, K. P. (2009). Assessing speech, language, and communication in autism spectrum disorders. In S. Goldstein, J. A. Naglieri, & S. Ozonoff (Eds.), *Assessment of autism spectrum disorders* (pp. 171–208). New York: Guilford Press.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. D., & Richardson, W. S. (1996, January). Evidence based medicine: What it is and what it isn't (Editorial). *British Medical Journal*, 13, 71–72. Retrieved from <http://www.bmj.com/cgi/content/full/312/7023/71>
- Schlosser, R. W., & Lee, D. L. (2000). Promoting generalization and maintenance in augmentative and alternative communication: A meta-analysis of 20 years of effectiveness research. *Augmentative and Alternative Communication*, 16, 208–226.
- Shavelson, R. J., & Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Snell, M. E., Chen, L. Y., & Hoover, K. (2006). Teaching augmentative and alternative communication to students with severe disabilities: A review of intervention research 1997–2003. *Research and Practice for Persons With Severe Disabilities*, 31, 203–214.
- Special Education Programs, Office of Special Education and Rehabilitation Services, USDE, and Technical Assistance Development System. (1985, August 19–21). *Children and youth with severe handicaps: Effective communication—A symposium*, Washington, DC.
- Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C, Paul, R., et al. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 52, 643–652.
- Torgerson, C. (2006). Publication bias: The Achilles' heel of systematic reviews? *British Journal of Education Studies*, 54, 89–102.
- Troia, G. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly*, 34, 28–53.
- Tuckman, B. W. (1999). *Conducting educational research* (5th ed.). Fort Worth, TX: Harcourt Brace.
- Yoder, P. J., & Layton, T. L. (1988). Speech following sign language training in autistic children with minimal verbal language. *Journal of Autism and Developmental Disorders*, 18, 217–229.

---

Received 9/1/2009, accepted 2/14/2010.

Editor-in-charge: Ann Kaiser

---

We thank Youngzie Lee, University of Virginia, and R. Michael Barker, Georgia State University, for their help with the analyses. We also thank the



National Center on Evidence-Based Practice in Communication Disorders of the American Speech-Language-Hearing Association for their assistance in conducting the systematic literature search and Dean Schofield, Appalachian State University, for his help in completing the literature search. The additional search terms, the NJC Evidence-Based Practices Data Entry

Instrument (2008), and the complete listing of studies meeting inclusion criteria are available on the NJC Web site: <http://professional.asha.org/njc/>. Correspondence regarding this article should be sent to Martha E. Snell, Curry School of Education, University of Virginia, PO Box 400273, Charlottesville, VA, 22904-4273. E-mail: [Snell@virginia.edu](mailto:Snell@virginia.edu)