Reading Instruction for Children who use AAC: Considerations in the Pursuit of Generalizable Results

R. Michael Barker, Kathryn J. Saunders, and Nancy C. Brady

R. Michael Barker, Schiefelbusch Institute for Life Span Studies, University of Kansas, United States; Kathryn J. Saunders, Schiefelbusch Institute for Life Span Studies at Parsons, University of Kansas, United States; Nancy C. Brady, Department of Speech-Language-Hearing: Sciences and Disorders, University of Kansas, United States

Abstract

Our purpose was to review evidence-based literacy instruction for children with severe speech impairment (SSI) who communicate with AAC. This review focuses on three issues important to researchers in this area: participant heterogeneity, assessment and instruction, and research design. We found 8 articles that reported attempts to teach phonological awareness and individual-word reading to a total of 26 children with SSI who used AAC. We evaluated these studies based on reporting of participant characteristics, assessment and instruction modifications, and the strength of research designs. We conclude by highlighting the need for standard assessments that can be used across studies, discussing strategies for facilitating metaanalyses, and suggesting the creation of an online database for researchers to share results on literacy instruction for this population.

Keywords

reading instruction; phonological awareness; AAC; single-subject design; assessment

The importance of adequate literacy skills cannot be overstated. Basic literacy proficiency is an important factor in determining job placement, medical health, and overall quality of life for people with typical development (Baer, Kutner, & Sabatini, 2009) and intellectual disabilities (Erickson, 2005). The 2003 National Assessment of Adult Literacy reported that at least 50% of people with typical development and below basic literacy skills—defined as being unable to answer questions that assess prose, document, and quantitative literacy—were unemployed or not in the labor force (Baer et al., 2009). Literacy acquisition is particularly important for children with severe speech impairments (SSI) who use augmentative and alternative communication (AAC). Reading and writing affords children and adults who use AAC the opportunity to communicate in a conventional, non-stigmatizing manner that is transparent to others in the community, particularly when used in conjunction with a speech-generating device. Unfortunately, people who use AAC have considerable difficulty acquiring literacy (Foley & Pollatsek, 1999; Koppenhaver & Yoder, 1992).

Little is known about the best approaches to literacy instruction for children who use AAC, as there are few published studies on the topic. Researchers and educators can leverage best instructional practices for the general population, however, with appropriate modifications for learners who do not speak, as the process of learning to read is theoretically similar for...
children who use AAC and children who are typically developing (Erickson, Koppenhaver, & Cunningham, 2006; Hetzroni, 2004).

Numerous studies have indicated that phonological awareness is the single largest predictor of reading outcomes (National Institute of Child Health and Human Development, 2000; Snow, Burns, & Griffin, 1998; Wagner & Torgesen, 1987; Wagner, Torgesen, & Rashotte, 1994). Phonological awareness refers to a person’s ability to focus on and manipulate phonemes in spoken syllables and words (NICHD, 2000). People who demonstrate high levels of phonological awareness—such as recognizing rhyming words or that certain words begin with the same sound—generally have an easier time learning to decode text. Decoding, in turn, is a fundamental reading skill that allows readers to “sound out” text to produce a spoken word. Because readers use decoding skills when they encounter unfamiliar words, reading instruction for children who use AAC should include components that support the development of phonological awareness and, ultimately, decoding skills.

The purpose of this paper is to review the existing literature on reading instruction for children with SSI who use AAC. To date, only one review has evaluated research on reading instruction for children who use AAC (i.e., Machalicek et al., 2010). The goal was to “guide and inform practitioners in evidence-based literacy instruction” (p. 221). These authors focused on studies that taught a broad range of participants—with physical and/or developmental disabilities, some with concomitant vision and hearing problems—using a broad range of instructional strategies with goals that ranged from improving phonological awareness and decoding to sight-word instruction and increasing symbolic communication. The present review took a closer look at specific component skills of decoding—including phonological awareness skills such as segmenting, blending, and letter-sound correspondence—with children who use AAC and have vision and hearing in the normal range. The major goals of this review are: a) to review the current literature on reading instruction for this population, b) to identify elements of the studies that are necessary for generalizable results, and c) to provide recommendations meant to help build a solid foundation for future research. To this end, we identified three challenges that researchers in this area face in the pursuit of generalizable results: describing participant characteristics, selecting and adapting assessment and instruction tasks, and selecting appropriate research designs.

**Participant Characteristics**

Given the heterogeneity of children with severe disabilities (Snell et al., 2010) and the single-subject research designs common to literacy instruction research with this population, it is especially important that research reports include explicit descriptions of several aspects of participants’ functioning. As Horner et al. (2005) notes, the description should be detailed enough for another researcher to select similar participants. Extant speech and reading skills are the most important to document because of their strong relationships with reading outcomes (Card & Dodd, 2006; Walley, Metsala, & Garlock, 2003). It is important that researchers assess these skills prior to intervention and report details of both procedures and results.

Adequately describing participants’ speech is important because of the relationship between articulation skills and phonological awareness. Children with typical speech tend to perform better on measures of phonological awareness than children with impaired speech; children with impaired speech tend to perform better than children with no speech (Card & Dodd, 2006; Foley & Pollatsek, 1999; Vandervelden & Siegel, 1999). Researchers have speculated that these differences could be due to difficulty with subvocal rehearsal in the phonological loop of working memory (Blischak, 1994; Card & Dodd, 2006). In addition, it is difficult to
measure phonological awareness in children who use AAC because many assessments rely on speech responses.

A detailed description of participants’ extant reading skills, such as letter–sound and name knowledge, identification of rhymes, and word identification and decoding skills, is also critical to replicability and interpretation of outcomes. Differences in these skills can be attributed to individual differences in speech skills, language, and history of reading instruction. Therefore, documenting participants’ extant skill levels can be critical to interpreting intervention outcomes within and across studies.

Assessment and Instruction

There are several assessment and instruction tasks used with children who speak. These include sound-matching, phoneme blending, letter–sound knowledge, word segmentation, spelling, and word identification. In a sound-matching task, such as the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) participants are given a spoken word, with an accompanying picture, and asked to identify the picture, from a set of choices, that starts with the same sound as the sample. In a phoneme blending task (e.g., CTOPP; Wagner et al., 1999), participants are presented with a series of phonemes that make up a spoken word. The participant blends those phonemes together and speaks the word. In assessments of letter–sound knowledge participants are presented with a printed letter and asked to produce the sound the letter makes (e.g., Lovett et al., 1994). In tasks that measure word segmentation (e.g., CTOPP; Wagner et al., 1999) participants are presented with a spoken word and asked to say the word one phoneme at a time. In spelling assessments, such as the Test of Written Spelling–4 (Larsen, Hammill, & Moats, 1999), participants are simply instructed to spell [target word]. Finally, word identification tasks assess real word reading (e.g., Woodcock, 1998). People are presented with a printed word and asked to read the word aloud.

For people who do not speak, it is necessary to modify the assessments that require spoken responses. Modifications include having participants point to (or otherwise indicate) items in an array of pictures, letters, or printed words. Such modifications, however, can fundamentally change the nature of the task. For example, the modified task provides a closed set of options and responses may be correct based on chance; if there are three choices, then 33% of responses would be correct even if the child guessed. This is not the case when participants answer using speech in an open answer format. Thus, standard scores or grade equivalents derived from the unmodified assessments (i.e., those presented in the examiner’s manual) cannot be used, as the tasks have been fundamentally changed. Moreover, researchers must describe any modifications made to assessments to facilitate readers’ interpretation of results.

Research Design

Single-subject designs are an important phase of conducting reading research with people who use AAC. Although large-scale clinical trials are widely accepted as the ultimate test of an instructional procedure (O’Toole, Logemann, & Baum, 1998), there are inherent difficulties conducting large-scale clinical trial studies with people who communicate with AAC. These difficulties include identifying sufficient numbers of people, randomly assigning people to treatment versus control conditions, and selecting an intervention that has sufficient preliminary data to warrant a clinical trial.

Large-scale clinical trials, however, are not the only means of determining the effectiveness of an intervention. Single-subject designs, such as alternating treatments and multiple-baseline designs, can provide strong evidence of intervention effects (Hersen & Barlow,
Alternating-treatments designs assess outcomes different conditions, or treatments, to determine which is most effective. Multiple-baseline designs replicate procedures across participants and/or behaviors or contexts. In multiple-baseline across-participants designs, performance is measured at the same intervals for all participants, but the intervention is introduced to each participant at a different time. An increase in accuracy only during the intervention phase indicates that the intervention is likely responsible for the increase.

As part of the What Works Clearinghouse (WWC), Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, et al. (2010) described minimum standards of evidence for single-subject designs. The WWC was established in 2002 by the Department of Education’s Institute of Education Sciences “to be a central and trusted source of scientific evidence for what works in education” (Institute of Education Sciences, 2012, About Us, para 2). Their minimum standards of evidence include “having at least three attempts to demonstrate an intervention effect at three different points in time or with three different phase repetitions” (p. 15). In addition, each phase should have a minimum of three data points. Finally, for data collected in observational contexts, two observers, one of whom is not informed of the underlying research question, should independently score at least 20% of the sessions and inter-observer agreement should be high.

Kratochwill et al. (2010) define three degrees of evidence for demonstrating that an intervention had an effect on outcomes. To be considered strong evidence, the study must include three demonstrations of the intervention effect (e.g., a change from baseline observation across participants) with no instances of non-effects (e.g., instances where there was no change from baseline). To demonstrate moderate evidence of an effect, the study must have at least three demonstrations of an effect, but can have one or more demonstrations of a non-effect. A study is said to have no evidence of an effect if it does not have at least three demonstrations of a change after introducing an intervention.

Overview of the Review

For this review we first summarize the characteristics of the participants who have been studied to date, and discuss the importance of describing extant communication skills, specifically speech and reading skills. Second, we describe tasks used to measure and teach reading and pre-reading skills in this population with a focus on adaptations to accommodate children with severe speech impairment who use AAC, and stress the importance of providing clear description of the measures used, including assessment, intervention, and outcome measures. Third, we discuss research designs that have been used and the importance of using strong designs by evaluating the studies using Kratochwill et al.’s (2010) criteria. We close by highlighting examples from the current literature of best research practices and discussing future directions for this field of study.

Method

The research reports chosen for this review were identified using the Google Scholar search engine (scholar.google.com). Google scholar searches the internet for journal articles, theses, dissertations, abstracts, and books contained in multiple databases across the internet from a single webpage (Google, 2010). This included the full catalog of holdings of the University of Kansas, numerous databases (e.g., Education Resources Information Center, EBSCOhost, PsycINFO, and MEDLINE), and journal publishers’ websites (e.g., Augmentative and Alternative Communication and Journal of Speech, Language, and Hearing Research). We searched for the following terms: “phonological awareness” or “phonological processing,” “reading” or “literacy,” and “augmentative and alternative communication” or “AAC” or “speech-generating device” or “SGD” or “voice output
communication aid” or “VOCA.” In addition, we investigated the reference lists of articles that fit our search criteria for articles that may not have been found in the online search.

In order to be included in this review, the articles had to meet the following criteria: (a) demonstrate a strategy for systematically teaching phonological awareness or decoding skills, (b) participants must have had a severe speech impairment, (c) at least one participant must have used either aided or unaided AAC as their primary mode of communication, (d) participants must have been described as having hearing and vision within normal limits (with correction if needed).

Results

We identified eight articles that fit our inclusion criteria (Blischak, Shah, Lombardino, & Chiarella, 2004; Coleman-Martin, Heller, Cihak, & Irvine, 2005; Fallon, Light, McNaughton, Drager, & Hammer, 2004; Hanser & Erickson, 2007; Heller, Fredrick, Tumlin, & Brineman, 2002; Johnston, Davenport, Kanarowski, Rhodehouse, & McDonnell, 2009; Millar, Light, & McNaughton, 2004; Truxler & O’Keefe, 2007). The studies all used single-subject designs with at least two, and up to five, participants. A summary of these articles can be found in Table 1. The following sections summarize how the articles relate to the three challenges previously outlined.

Participant Characteristics

A total of 26 people with SSI who use AAC participated in the eight studies. Sixteen used speech-generating devices, six used communication boards without voice-output, and four used unaided AAC systems to augment their communication. One was an adult (CA = 23); the rest ranged in age from 4 to 16 years old. Nineteen participants had an intellectual disability. Sixteen were diagnosed with cerebral palsy. Table 1 indicates which participants had an intellectual disability and, if they had another medical diagnosis, what that diagnosis was. Intellectual disability and/or cerebral palsy were the most common precipitating factors for the participants’ speech difficulties.

Extant speech skills—One article was particularly detailed in its description of speech intelligibility. Fallon et al. (2004) reported that their five participants scored 10% to 30% on a measure of the percentage of intelligible words produced using standardized stimulus words from the Index of Augmented Speech Comprehensibility in Children (Dowden, 1997). This assessment provided a pool of 300 different words with corresponding pictures from which 30 were randomly chosen for administration. Participants’ picture-naming responses were audio recorded, and unfamiliar listeners transcribed the recordings without seeing the pictures. Percent intelligibility was determined by dividing the number of words transcribed correctly by the total number of items. Thus, the authors clearly described the stimuli used, the method for recording participant responses, and the criteria used for intelligibility judgments.

Each of the remaining studies reported information about participants’ speech, but with less detail. Millar et al. (2004) reported that their participants’ speech was less than 50% intelligible to an unfamiliar listener. Blischak et al. (2004) reported that participants’ speech intelligibility ranged from 20% to 35% for unfamiliar listeners. Johnston et al. (2009) reported that their two participants spoke five words that were intelligible to familiar listeners and could say yes and no. Truxler and O’Keefe (2007) reported that their participants’ speech was “limited to 1- to 2-word utterances that were intelligible to familiar listeners” (p. 165). Hanser and Erickson (2007) stated that their participants could not use speech to meet face-to-face communication needs. Importantly, none of these studies described the procedures used to assess intelligibility, hence limiting replicability and
across-study comparisons. The remaining studies provided a description of speech by reporting clinical diagnoses, such as anarthria or dysarthria, without a specific measure of speech intelligibility (Coleman-Martin et al., 2005; Heller et al., 2002).

**Extant reading skills**—Three articles used experimenter-designed measures of knowledge about the relationships between letters, their names, and the sounds they represent, prior to instruction. Millar et al. (2004) reported letter-identification skills and initial-sound identification scores. Their three participants identified the correct letter, from its spoken name in an array of four letters, with 88%, 96%, and 70% accuracy. They identified the first letter of a spoken word from a field of four letters with 6%, 10%, and 0% accuracy, respectively. Millar et al. recruited participants who could recognize at least 70% of the letters of the alphabet. Similarly, Blischak et al. (2004) reported that their participants recognized the 10 target letters used in the intervention (i.e., d, h, k, l, m, n, p, a, o, u) by name and that they scored 30% accurate or less in a measure of sound–letter correspondence for the same letters. Fallon et al. (2004) reported that participants recognized a minimum of 50% of the target letters used in their intervention (i.e., p, b, n, t, g, s, l, r, e, i, a, o, u) and a minimum of 50% of the sounds that corresponded with the same letters. Also, they reported that their participants were taught, prior to the intervention, to identify 100% of these letters by name and 100% of the sounds that corresponded to the letters.

Four other articles reported information about participants’ reading skills using either standardized assessments with modified administration, experimenter-designed assessments, or teacher checklists. Heller et al. (2002) reported that the three participants in their study scored a grade equivalent of 2.0, 2.25, and 1.5 on the word identification subscale of the Woodcock Reading Mastery Test (WRMT), the Peabody Individual Achievement Test, and the WRMT, respectively. Likewise, Coleman-Martin et al. (2005) reported that the three participants in their study had grade-equivalent reading scores of 2.9, 2, and 1 according to the Peabody Individual Achievement Test-Revised (PIAT-R; Markwardt, 1989), an informal teacher assessment that was not described, and an assessment that was not described, respectively. Truxler and O’Keefe (2007) stated that their participants demonstrated scores at or below 40% correct for a modified administration of the Woodcock Reading Mastery Test – Revised (Woodcock, 1998) word identification and word attack subscales. It was unclear, however, how many items were administered from the Woodcock. Heller et al., Coleman-Martin et al., and Truxler and O’Keefe do not describe the modifications were made to these standardized assessments even though they were administered to children with severely limited speech who used AAC. Hanser and Erickson (2007) reported that their three participants could identify 14, 17, and 22 out of 25 on a word-recognition task, although the details of task administration and specific distractor items were not given. Finally, Johnston et al. (2009) demonstrated that the two participants in their study demonstrated symbolic knowledge, some print awareness, and some letter identification skills, according to the Ladders to Literacy preschool checklist (Notari-Syverson, O’Connor, & Vadasy, 1998).

**Assessment and Instruction**

Many different phonological awareness and reading assessments and instruction strategies were used in the articles we reviewed. Table 2 contains a summary of the tasks used in these studies, along with the modifications that made them appropriate for children who use AAC.

**Sound-matching**—Two studies used sound-matching tasks for assessment and instruction (i.e., Fallon et al., 2004; Truxler & O’Keefe, 2007). For example, Fallon et al. (2004) used a sound-matching task that was conceptually similar to the task in the CTOPP (Wagner et al., 1999). They presented participants with four pictures of known items, each starting with a
different initial phoneme, and spoke the names of each item. They then presented a single phoneme, spoke the names of the items again, and asked which of the pictures started with the single phoneme.

**Phoneme blending**—Four studies used phoneme blending for assessment and/or instruction (i.e., Coleman-Martin et al., 2005; Fallon et al., 2004; Heller et al., 2002; Truxler & O'Keefe, 2007). Typical of the modifications made in these studies, Fallon et al. (2004) eliminated the speech requirement and had participants choose a picture that represented the target word. For each item, participants were presented with four pictures: one target and three foils. Of the three foils, one contained a different initial sound, one contained a different vowel, and one contained a different final sound.

**Letter–sound knowledge**—Six studies used modified letter–sound knowledge tasks for either assessment or instruction (i.e., Blischak et al., 2004; Fallon et al., 2004; Hanser & Erickson, 2007; Johnston et al., 2009; Millar et al., 2004; Truxler & O'Keefe, 2007). For children who use AAC, it is necessary to provide the phoneme and have the child choose the letter. For example, Johnston et al. (2009) presented participants with arrays of eight printed letters (i.e., a, m, t, s, i, f, d, r). Participants were then asked to “touch the letter that says [target sound]” (Johnston et al., 2009, p. 127). The participants touched the letter that corresponded to the spoken phoneme.

**Word segmentation**—Six studies used variations on word segmentation tasks for assessment and/or instruction (i.e., Blischak et al., 2004; Coleman-Martin et al., 2005; Fallon et al., 2004; Heller et al., 2002; Millar et al., 2004; Truxler & O'Keefe, 2007). Three studies used initial phoneme segmentation tasks (Fallon et al., 2004; Millar et al., 2004; Truxler & O'Keefe, 2007). For example, in Millar et al. (2004), participants were presented with a spoken word that was already in their receptive vocabulary. They were then asked to select the letter the spoken word began with on an adaptive keyboard. In order to complete this task, participants had to segment the initial sound from the rest of the word, and then choose the letter that corresponded to the sound. In addition, one study taught participants to segment entire words into individual phonemes by using checkers to mark the number of phonemes in the word (Blischak et al., 2004). The two remaining studies taught segmenting by explicitly pointing out how printed letters mapped onto the phonemes of the word, as the word was being pronounced one phoneme at a time (Coleman-Martin et al., 2005; Heller et al., 2002).

**Spelling**—Four studies used spelling tasks for assessment and/or instruction (i.e., Blischak et al., 2004; Hanser & Erickson, 2007; Johnston et al., 2009; Truxler & O'Keefe, 2007) that did not use a paper and pencil response mode. Blischak et al. (2004) had participants spell by arranging tiles with printed letters. Hanser and Erickson (2007) had participants spell by selecting letters on their AAC devices. The participants in the Johnston et al. (2009) study spelled by pointing in sequence to letters printed on a piece of paper. Truxler and O’Keefe (2007) had participants spell by selecting letters on a large cardboard qwerty keyboard.

**Word identification**—Five articles used variations of word identification tasks for assessment and/or instruction (e.g., Coleman-Martin et al., 2005; Fallon et al., 2004; Hanser & Erickson, 2007; Heller et al., 2002; Truxler & O'Keefe, 2007). Three studies used a printed word as a sample (Coleman-Martin et al., 2005; Fallon et al., 2004; Heller et al., 2002). Fallon et al. (2004) modified the assessment so that participants chose pictures to indicate they had read a printed word correctly. They presented participants with a printed word, highlighted in yellow, on a page with four line-drawn pictures at the bottom. The participants were asked to read the word and indicate, by pointing, which picture the printed
Coleman-Martin et al. (2005) and Heller et al. (2002) used spoken words, instead of pictures, as response choices. This strategy has the added advantage of assessing participants’ ability to decode nonwords as well as real words, because nonwords typically cannot be pictured. For example, the researcher presents the printed word diz and asks, does that spell cat (the child responds yes or no)? Dog (yes/no)? Diz (yes/no)?

Two studies used a spoken word as the sample, instead of a printed word, allowing responses to be made by pointing to a printed word in an array (Hanser & Erickson, 2007; Truxler & O’Keefe, 2007). For example, Truxler and O’Keefe (2007) presented their participants with a spoken target word and then asked them to choose the printed word, from an array of 10, that matched the spoken word.

**Teaching Approaches**—Five articles included instruction that directly and explicitly taught phonological awareness, reading, and/or spelling skills (Blischak et al., 2004; Fallon et al., 2004; Johnston et al., 2009; Millar et al., 2004; Truxler & O’Keefe, 2007). For example, for phoneme blending instruction, Fallon et al. (2004) presented participants with three phonemes spoken by the instructor (e.g., /m/ – /o/ – /p/). The participant was then asked to point to the picture of the word that those sounds made. If the participant did not answer correctly, the instructor provided an explanation of the correct response. Next, the instructor led the participant in making the correct response. Finally, the instructor allowed the participant an opportunity to respond independently. Similarly, Millar et al. (2004) used a most-to-least prompting hierarchy when participants made errors in identifying initial letters of spoken words. For the full prompt, the instructor spoke the word and elongated the first sound and placed a space between the initial sound and the rest of the word. For the partial prompt, the instructor only stressed and elongated the initial sound. For the no-prompt level, the instructor said the word normally. Johnston et al. (2009) used a different prompting strategy in their instruction for sound-letter correspondence and spelling of CVC words. For the initial sessions, instructors pointed to letters as they presented the item (either phoneme or spoken word). Once participants achieved a criterion of 80% correct for the prompted items, the instructors introduced a 5-second delay between the presentation of the item and prompt to allow the participant time to respond without the model.

Three articles used two previously published intervention strategies that were designed for children who use AAC, both of which included direct instruction components. Heller et al. (2002) and Coleman-Martin et al. (2005) used the *Nonverbal Reading Approach* (Heller, Fredrick, & Diggs, 1999) to teach word identification. For the first presentation of a printed word, the instructor sounded out the word along with the student (who was instructed to use internal speech or speech approximations). Next, the instructor pointed to each of the letters of the printed word and spoke the phoneme represented by the letter. The participant was instructed to follow along. Next, the instructor covered all but one of the letters and said the sound made by the uncovered letter. The participant was encouraged to say the sound in his or her head. Next, the instructor (out loud) and participant (silently, “in his or her head”) said the sound together, in succession. Then they said the word quickly, blending the sounds together to construct the target word (again, the participant presumably did so silently).

Hanser and Erickson (2007) used the *Literacy Through Unity: Word Study* (Erickson & Hanser, 2007) program. This program was unique in the instructional strategies we reviewed because it was integrated into the AAC systems that the participants used. The program contained three groups of scripted lessons: word wall, making words with letters, and making words with icons, each with 25 lessons. The word wall lessons supported learning high frequency words in an automatic fashion and focused on using knowledge of those words to read other words with the same spelling patterns. Making words with letters consisted of phonological awareness instruction in both reading and spelling. Making words...
with icons supported combining the symbols used in the Unity™ system into sequences to convey different meanings. The emphasis of this program was on the generalization of phonological and decoding skills to unfamiliar literacy situations.

Research Design

Each study reviewed used a single-subject research design. We evaluated each article to determine whether it met Kratochwill et al.’s (2010) minimum evidence standards, and if so, whether it demonstrated strong, moderate, or no evidence of an effect. These results are summarized in Table 1. Two studies we reviewed did not meet minimum evidence standards (Blischak et al., 2004; Hanser & Erickson, 2007). For Blischak et al. (2004), it is unclear whether reliability data were taken in 20% of the sessions. Hanser and Erickson (2007) did not provide sufficient numbers of data points. Six studies met minimum evidence standards. Coleman-Martin et al. (2005) did so using a modified alternating-treatments design. The remaining articles that met minimum standards criteria (Fallon et al., 2004; Heller et al., 2002; Johnston et al., 2009; Millar et al., 2004; Truxler & O’Keefe, 2007) used multiple-baseline across-participants designs. These articles had high inter-rater agreement for observations and at least three data points for each instruction phase. Consequently, these articles were evaluated for the strength of the effects of reading instruction.

Of those six studies, Millar et al. (2004) demonstrated no evidence of an effect because there were not at least three positive demonstrations (only 2 of their 3 participants increased in initial phoneme segmentation and letter–sound knowledge following intervention). The remaining five studies met criteria for strong evidence for the effect of instruction on outcomes. Coleman-Martin et al. (2005), who used an alternating treatments design, demonstrated increases in the correct selection of printed words given a spoken-word sample following the introduction of teaching conditions across three different sets of words for each of three participants. The remaining four studies (Fallon et al., 2004; Heller et al., 2002; Johnston et al., 2009; Truxler & O’Keefe, 2007), all of which used multiple-baseline-across-participants designs, demonstrated three or more changes from baseline across participants. Heller et al. (2002) documented large increases in word reading measures following the introduction of two different instruction phases in three children. Johnston et al. (2009) documented large increases in letter–sound knowledge for three letters and large increases in spelling CVC words using the same letters following the introduction of a three-step instruction program in two children. Truxler and O’Keefe (2007) documented small increases in initial-sound segmentation following the introduction of instruction in four children. Finally, Fallon et al. (2004) documented large increases in word reading following sound-matching, blending, and word identification instruction in five children.

The effects discussed to this point were shown with words and sounds that were directly taught. To demonstrate that phonological awareness or decoding skills have been taught, however, it is necessary to show generalization to words that are different than those used in instruction. For example, phonological awareness is demonstrated when a child decodes or spells words that were not explicitly taught. Of the six studies that met minimum criteria for evaluation, five measured generalization of reading or spelling to untaught words (Coleman-Martin et al., 2005; Fallon et al., 2004; Heller et al., 2002; Johnston et al., 2009; Truxler & O’Keefe, 2007). Heller et al. (2002), Johnston et al. (2009), and Truxler and O’Keefe (2007) did not include at least three measurement occasions and, thus, did not meet minimum evidence standards for the generalization phase. Coleman-Martin et al. (2005) and Fallon et al. (2004) did meet minimum evidence standards for the generalization phase. The participants in Coleman-Martin et al. did not show any evidence of generalization of the reading skills taught during instruction to untaught words. Fallon et al. provided the most compelling demonstration of generalization, although the generalization data met criteria for only moderate evidence of the effect of instruction on generalization trials (because 2 of 5
participants did not demonstrate generalization). Fallon et al. taught five participants to decode words using intensive phonological awareness instruction, as mentioned previously. After participants achieved criterion with target words, the authors tested whether participants could decode untaught words that contained the same letters and phonemes. Results indicated that three participants read more of the novel words in a reading probe following intervention than they did at baseline. Thus, Fallon et al. provided compelling evidence that some participants in their study learned the underlying phonological awareness skills and that they could use these skills to decode in singleword reading contexts.

**Discussion**

The body of research on reading instruction and intervention for children who use AAC is small, but continuing to grow. This literature is small, in part, because it is conducted with a low prevalence population. This highlights the importance of maintaining high quality research standards when working with this population by providing descriptions of participants’ characteristics prior to instruction, descriptions of experimenter designed or modified standardized assessments, and using strong research designs.

Each of the studies we reviewed included several of these important characteristics. For example, all of the studies described participants’ speech intelligibility, although Fallon et al. (2004) did so with the most detailed description and objective procedure. Likewise, each study mentioned extant reading skills, but only a few of those using experimenter-designed reading measures provided detailed descriptions of how those measures were constructed and administered (i.e., Blischak et al., 2004; Fallon et al., 2004; Millar et al., 2004). Consequently, it is difficult to build a cumulative knowledge base because of the wide variety of information and assessments reported. For example, no two studies used the same procedure to assess intelligibility, and numerous measures of early literacy skills were used across studies. Consistent and detailed reporting of measures that have been agreed upon in the field would help build a needed, coherent, and cumulative knowledge base.

A strong suit of this group of studies was the use of experimental single-subject designs; 7 of the 8 studies did so (Blischak et al., 2004; Coleman-Martin et al., 2005; Fallon et al., 2004; Heller et al., 2002; Johnston et al., 2009; Millar et al., 2004; Truxler & O’Keefe, 2007). In addition, four articles included maintenance and generalization phases or items (Blischak et al., 2004; Fallon et al., 2004; Johnston et al., 2009; Millar et al., 2004). The study by Fallon et al. (2004) was the only one we reviewed that included each of the components described previously, met the minimum standards of evidence established by the WWC, provided strong evidence of intervention effects, and moderate evidence of generalization. Their five participants had steady baselines and demonstrated increased word reading only after intervention began. Furthermore, they did this in both reading probes and in book-reading contexts. Finally, 3 of the 5 participants demonstrated generalization by reading novel words composed of letters targeted in the intervention. In fact, the results of this study, along with those of Millar et al. (2004) were used to develop a literacy curriculum designed for children who use AAC (Light, McNaughton, Weyer, & Karg, 2008). The Accessible Literacy Learning curriculum (Light & McNaughton, 2009) incorporates the instructional strategies and materials shown to be effective in these studies and is therefore a good example of research informing practice.

**Future Directions**

There is much work to do in the area of reading intervention for children who use AAC. First, there is a strong need for standardized assessments of both phonological awareness and reading that do not require speech responses. Although we have discussed several modifications that can be made to phonological awareness and reading assessments, there...
have been few attempts to assess the validity of these new assessments by making comparisons to established assessments.

The importance of valid assessments becomes apparent when considering differences in task requirements among three word-reading tasks (i.e., the unmodified task and two modified tasks that do not require speech). The unmodified task requires the discrimination of printed words that are presented one by one, and does not constrain responses to a small number of choices. Neither modified task has both of these characteristics. One modified task involves a spoken word as the sample and printed words as choices (e.g., Truxler & O'Keefe, 2007). Like unmodified reading, correct responses demonstrate relations between speech and print. Unlike reading, correct responding might depend on the presence of the printed-word choices or even the specific printed words that are used as incorrect choices. A second modified task involves a printed-word sample with pictures as choices (e.g., Fallon et al., 2004). Like reading, this task requires the discrimination of printed words presented one by one. The spoken word is not directly involved in the task, however, thus this is not a direct measure of relations between speech and sound. These points are not meant as criticisms of these procedures, as speech responses are not possible. Consequently, it may be helpful to use both modified reading tasks, as together the two tasks would include more components of the unmodified word-reading task.

Similar considerations of validity arise when making other modifications described previously. For example, the modified phoneme blending task differs from the unmodified CTOPP (Wagner et al., 1999) in that selecting the picture requires the comprehension of the spoken word, and also that the child’s name for a picture must be the same as that of the developers of the test. Likewise, when having children choose letters from an array for a letter–sound knowledge task, instead of producing the phoneme, selections can be correct by chance. Furthermore, using an array of answer choices—be it pictures, letters, or printed words—allows children to compare correct and incorrect choices. This is not possible when individual printed words are presented, and thus fundamentally changes the nature of the task.

Researchers have attempted to validate some assessments for people who use AAC. One, the Assessment of Phonological Awareness and Reading (APAR; Iacono & Cupples, 2002; 2004), is available online in print and digital formats. The APAR measures reading using word recognition, nonword recognition, and word comprehension subscales, and phonological awareness using phoneme blending, phoneme counting, and sound-matching subscales. Iacono and Cupples (2004) found evidence of construct validity of the APAR in a study of 40 adults with significant disabilities. The APAR, however, has not been widely adopted (see Hart, Scherz, Apel, & Hodson, 2007, for one example) and its validity for use with children has yet to be established. In a second example, Gillam, Fargo, Foley, and Olszewski (2011) published validity data from children with typical development in dynamic phoneme-deletion task that used pictures as a response mode. Results indicated high construct validity, as performance on the dynamic phoneme deletion task was highly correlated to a deletion task that required speech responses and to other reading tasks. These studies represent important first steps in developing valid assessments with uniform stimuli for children who use AAC.

Second, there is a strong need for replication of results to bolster external validity. The single-subject studies that were reviewed here answered questions about whether, and to what extent, specific instruction affected phonological awareness and reading skills for the study participants (Horner et al., 2005). But what is the generality of these outcomes to other participants? Single-subject designs derive their strength from replication. Ultimately, a research program aimed at the development of instruction seeks to determine not only
whether the instruction is effective, but also to determine the skills that the learner must already have to benefit from the instruction. Unfortunately, the literature currently contains so few replications of any given procedure that the generality of the outcomes across participants with different characteristics has not been established.

One way to address this issue is to compile the results of single subject-design studies using meta-analysis strategies. Shadish, Rindskopf, and Hedges (2008) and Kratochwill et al. (2010) discuss strategies for such meta-analyses, a full discussion of which is outside the scope of this review. They offer the following recommendations for researchers using single-subject designs. First, +unstandardized outcomes (i.e., raw scores) using the same metric and on the same scale are preferred because of the ease at which these can be compared within and across studies. Second, when unstandardized scores on the same scale are not available, regression based approaches to calculating standard scores are preferable (see Kratochwill et al., 2010, for a discussion). Summaries across cases and across studies, such as means and standard deviations, can be calculated when common metrics are used. The potential for summarizing data highlights the importance of common measures. These summaries would allow researchers to make judgments, based on effect sizes, about which interventions are the most effective for children with particular profiles. Ultimately, what is learned from these analyses can be used to inform large-scale intervention studies with a focus on generalizability to the larger population of children who use AAC. Doing this would require large multisite research programs similar to those used in research on interventions for reading disabilities in children who are typically developing (e.g., Morris et al., 2010).

Finally, the field of reading research with children who use AAC would benefit from the formation of an online database to aggregate results, similar to those used by researchers and clinicians in other research areas, such as the National Outcomes Measurement System (NOMS; American Speech-Language-Hearing Association) and the Child Language Data Exchange System (CHILDES; MacWhinney, 1996). Both databases allow researchers and clinicians to submit data that can be used by others to answer broad research questions. Because of its emphasis on intervention, the NOMS database can help researchers and clinicians answer questions ranging from which intervention approaches are appropriate for individual children to larger questions about which interventions are most effective overall. Another strong suit of NOMS is that all who submit data to the database are required to use a set of uniform measures. This enables many data points to be aggregated easily so that the overall effectiveness of particular intervention strategies can be evaluated. The availability of a similar database for reading research with children who use AAC would beneficial for clinicians, educators, and researchers alike.

Conclusion

Currently the evidence base for reading interventions for children who use AAC consists of a small group of single-subject-design studies that use different measures and teaching strategies and demonstrate varying degrees of success. There remains a strong need to develop standard phonological awareness and reading assessments with agreed upon sets of items, both in terms of targets and distractors, to answer questions about the effectiveness of interventions. There also is a need to replicate the results of effective reading interventions across different contexts and groups of children, in single-subject studies and ultimately in larger multisite clinical trials. These represent important and necessary steps in creating an empirically supported evidence base for reading instruction for children who use AAC.
Acknowledgments

This review was written with the support of NIH grants T32 HD057844, P30 HD002528, P01 HD018955, R01 DC007684, and R01 HD048528, awarded to the University of Kansas.

References


Larsen, SC.; Hammill, DD.; Moats, L. Test of written spelling. 4th ed.. Austin, TX: Pro-Ed; 1999.


Wagner, RK.; Torgesen, JK.; Rashotte, CA. Comprehensive test of phonological processing. Austin, TX: Pro-Ed, Inc; 1999.


## Table 1

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Child</th>
<th>Age</th>
<th>Diagnosis</th>
<th>IDD</th>
<th>AAC</th>
<th>Speech Intelligibility</th>
<th>Pre-Intervention Reading</th>
<th>MES?</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blischak et al. (2004)</td>
<td>Gabe</td>
<td>7;0</td>
<td>SLD</td>
<td>No</td>
<td>Unaided</td>
<td>20–30%</td>
<td>Phoneme-Grapheme correspondence &gt; 50%. ID 10 target letters.</td>
<td>No</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Larry</td>
<td>5;0</td>
<td>SLD</td>
<td>No</td>
<td>SGD</td>
<td>30–35%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jane</td>
<td>6;2</td>
<td>SLD</td>
<td>No</td>
<td>Unaided</td>
<td>20–30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coleman-Martin et al. (2005)</td>
<td>Alice</td>
<td>11;0</td>
<td>CP</td>
<td>No</td>
<td>SGD</td>
<td>Severe Speech Impairment</td>
<td>Demonstrated grapheme-phoneme correspondence; 1st to 3rd grade reading level</td>
<td>Yes</td>
<td>Strong Evidence</td>
</tr>
<tr>
<td></td>
<td>Beth</td>
<td>12;0</td>
<td>Autism</td>
<td>Mod</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carrie</td>
<td>16;0</td>
<td>TBI</td>
<td>Mod</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fallon et al. (2004)</td>
<td>Dale</td>
<td>11;9</td>
<td>IDD</td>
<td>Mod</td>
<td>SGD</td>
<td>25%</td>
<td>ID 50% of target letters when named; 100% after pre-training.</td>
<td>Yes</td>
<td>Strong Evidence</td>
</tr>
<tr>
<td></td>
<td>Sam</td>
<td>14;0</td>
<td>IDD</td>
<td>MM</td>
<td>SGD</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tommy</td>
<td>9;11</td>
<td>CP</td>
<td>Mod</td>
<td>Unaided</td>
<td>30%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amy</td>
<td>10;3</td>
<td>CP</td>
<td>No</td>
<td>Unaided</td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nate</td>
<td>9;5</td>
<td>DS</td>
<td>Mod</td>
<td>SGD</td>
<td>10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanser &amp; Erickson (2007)</td>
<td>Child 1</td>
<td>13;2</td>
<td>CP</td>
<td>No</td>
<td>SGD</td>
<td>Unable to use speech to meet face-to-face communication needs</td>
<td>ID 37 or more capital and lowercase letters; score 11/13 on concepts of print; high Word ID scores</td>
<td>No</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Child 2</td>
<td>13;1</td>
<td>CP</td>
<td>Mild</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child 3</td>
<td>7;2</td>
<td>CP</td>
<td>No</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heller et al. (2002)</td>
<td>Al</td>
<td>9;0</td>
<td>HOS</td>
<td>Mild</td>
<td>SGD</td>
<td>Anarthria</td>
<td>Demonstrated grapheme-phoneme correspondence; 1st to 2nd grade Word ID</td>
<td>Yes</td>
<td>Strong Evidence</td>
</tr>
<tr>
<td></td>
<td>Betsy</td>
<td>16;0</td>
<td>CP</td>
<td>Mild</td>
<td>SGD</td>
<td>Anarthria, Severe Dysarthria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cathy</td>
<td>23;0</td>
<td>CP</td>
<td>Mild</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnston et al. (2009)</td>
<td>Eddie</td>
<td>5;3</td>
<td>CP</td>
<td>Mild</td>
<td>CB</td>
<td>5 words including “yes” and “no”</td>
<td>Symbolic knowledge, some print awareness.</td>
<td>Yes</td>
<td>Strong Evidence</td>
</tr>
<tr>
<td></td>
<td>Mary</td>
<td>4;2</td>
<td>CP</td>
<td>Mild</td>
<td>CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Millar et al. (2004)</td>
<td>Melinda</td>
<td>7;0</td>
<td>CP</td>
<td>No</td>
<td>SGD</td>
<td>Less than 50%</td>
<td>ID 70% of letters by name. Each could ID some words by sight.</td>
<td>Yes</td>
<td>No Evidence</td>
</tr>
<tr>
<td></td>
<td>Haley</td>
<td>10;0</td>
<td>CP</td>
<td>Mod</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gary</td>
<td>10;0</td>
<td>CP</td>
<td>Mod</td>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truxler &amp; O’Keefe (2007)</td>
<td>Child 1</td>
<td>8;0</td>
<td>CP</td>
<td>DNS</td>
<td>CB</td>
<td>1 to 2 word utterances intelligible to familiar listeners</td>
<td>40% or less correct on modified Word ID and Word Attack</td>
<td>Yes</td>
<td>Strong Evidence</td>
</tr>
<tr>
<td></td>
<td>Child 2</td>
<td>9;6</td>
<td>CP</td>
<td>DNS</td>
<td>CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child 3</td>
<td>8;7</td>
<td>CP</td>
<td>DNS</td>
<td>CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child 4</td>
<td>9;6</td>
<td>CP</td>
<td>DNS</td>
<td>CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Age in years;months. SLD = Speech and/or Language disorder; CP = Cerebral Palsy; TBI = Traumatic Brain Injury; IDD = Intellectual and Developmental Disability; DS = Down syndrome; HOS = Holt-Oram Syndrome; Mod = Moderate; MM = Mild to Moderate; DNS = Yes, did not specify severity; SGD = Speech-generating device; CB = non-electronic communication board; ID = Identification; MES? = Met minimum evidence standards?
Assessment Modifications for Children who use AAC

<table>
<thead>
<tr>
<th>Unmodified Assessment</th>
<th>Stimulus</th>
<th>Task</th>
<th>Standard Response</th>
<th>Modification</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound-matching; Sound Categorization;</td>
<td>Spoken Word (sometimes</td>
<td>Identify sound in given location (e.g.,</td>
<td>Respond by pointing to picture only</td>
<td>Fallon et al., 2004;</td>
<td></td>
</tr>
<tr>
<td>Phoneme Identity</td>
<td>presented with picture)</td>
<td></td>
<td></td>
<td>Truxler &amp; O’Keeffe, 2007</td>
<td></td>
</tr>
<tr>
<td>Phoneme Blending</td>
<td>Individual phonemes spoken</td>
<td>Blend together to generate word</td>
<td>Respond by pointing to picture of created word</td>
<td>Fallon et al., 2004;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in isolation</td>
<td></td>
<td></td>
<td>Truxler &amp; O’Keeffe, 2007</td>
<td></td>
</tr>
<tr>
<td>Letter–Sound Knowledge; Grapheme–Phoneme</td>
<td>Printed Letter</td>
<td>Recall phoneme</td>
<td>Present phoneme, respond by pointing to printed letter</td>
<td>Fallon et al., 2004;</td>
<td></td>
</tr>
<tr>
<td>Correspondence</td>
<td></td>
<td></td>
<td></td>
<td>Johnston et al., 2009;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Millar et al, 2004</td>
<td></td>
</tr>
<tr>
<td>Word Segmentation</td>
<td>Spoken Word</td>
<td>Break word into constituent phonemes</td>
<td>Place marker on table to indicate each phoneme</td>
<td>Blischak et al., 2004</td>
<td></td>
</tr>
<tr>
<td>Initial Phoneme Segmentation; Onset</td>
<td>Spoken Word</td>
<td>Isolate initial phoneme</td>
<td>Respond by pointing to printed letter that corresponds</td>
<td>Fallon et al., 2004;</td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td></td>
<td></td>
<td>to phoneme</td>
<td>Millar et al, 2004</td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>Spoken Word</td>
<td>Generate letters that correspond to phonemes in spoken word</td>
<td>Use tiles printed with letters or type on modified or standard keyboard</td>
<td>Blischak et al., 2004;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Johnston et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Word Identification</td>
<td>Printed Word</td>
<td>Read printed word</td>
<td>Respond by pointing to a picture that corresponds to</td>
<td>Fallon et al., 2004;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>the printed word.</td>
<td>Hamer &amp; Erickson, 2007</td>
<td></td>
</tr>
</tbody>
</table>