

*EQUIVALENCE RELATIONS IN INDIVIDUALS WITH LANGUAGE LIMITATIONS AND MENTAL RETARDATION*

JENNIFER O'DONNELL AND KATHRYN J. SAUNDERS

THE UNIVERSITY OF KANSAS

The study of equivalence relations exhibited by individuals with mental retardation and language limitations holds the promise of providing information of both theoretical and practical significance. We reviewed the equivalence literature with this population, defined in terms of subjects having moderate, severe, or profound mental retardation. The literature includes 55 such individuals, most of whom showed positive outcomes on equivalence tests. The results to date suggest that naming skills are not necessary for positive equivalence test outcomes. Thus far, however, relatively few subjects with minimal language have been studied. Moreover, we suggest that the scientific contributions of studies in this area would be enhanced with better documentation of language skills and other subject characteristics. With recent advances in laboratory procedures for establishing the baseline performances necessary for equivalence tests, this research area is poised for rapid growth.

*Key words:* stimulus equivalence, mental retardation, language limitations, relational learning, conditional discrimination, humans

The 30th anniversary of the publication of Sidman's first paper on equivalence relations was marked in 2001. The study (Sidman, 1971) demonstrated rudimentary reading comprehension of 20 words in an adult male with severe mental retardation (MR). At the beginning of the study, the man could neither match the printed words to representational pictures nor select the printed words upon hearing them spoken. He could, however, select the corresponding pictures upon hearing most of the words spoken, and he was taught to select the pictures that he did not initially know. He also could correctly name most of the pictures before any training. He was taught to select the corresponding printed word upon hearing each of the 20 words spoken. After mastering the spoken-to-printed word task, the subject then selected printed words to match their corresponding pictures, and vice versa. Further, he named most of the printed words. This outcome demonstrated that the man learned more about the stimuli than was taught di-

rectly. Emergent performances like these are critical components of language development.

Sidman and Tailby (1982) formalized a set of tests designed to distinguish conditional relations that were merely unidirectional, rote chains from those that could serve as prerequisites of emergent relations. They called the latter "equivalence relations." Sidman and colleagues proposed that the definition of equivalence relations captured essential properties of the kind of relational learning that is typically labeled as symbolic or representational. For example, consider a child who learns to select the numeral 2 (from among several numerals) when he hears the word "two," and also learns to select the printed word two (from among several words) when he sees the numeral 2. In a positive test for reflexivity, the child selects the numeral 2 when shown the numeral 2. In a test for symmetry, the printed word is presented as the sample, and the child must select from one or more numeral choices. In other words, the sample and comparison stimuli change places. In the test for transitivity, the spoken word sample is presented with printed word choices. The foundation for this relation is that the spoken and the printed word are related to the same picture. A conditional relation that has all three properties—reflexivity, symmetry, and transitivity—is called an equivalence relation. It is also possible to present a combined test for sym-

---

The first author was supported by an NICHD Institutional National Research Service Award (T32 HD07525) awarded to the University of Kansas Institute for Life Span Studies. Preparation was also supported by the program project "Communication of People with Mental Retardation" (PO1 HD18955). Jennifer O'Donnell is now at Allegheny College. We thank Joe Spradlin for his helpful comments and Pat White and Julie McEntee for their editorial assistance.

Send reprint requests to Kathryn Saunders, Parsons Research Center, 2601 Gabriel, Parsons, Kansas 67357 (e-mail: ksaunders@ku.edu).

metry and transitivity. As in the test for transitivity, this test is for emergent relations between two stimuli that are related to a common third stimulus. As in the test for symmetry, a stimulus that has not previously served as a sample does so. This combined test has been referred to as a test for equivalence (Sidman & Tailby).

The publication of Sidman and Tailby (1982) led to the rapid growth of the research on equivalence relations. Despite 30 years of research on this topic, however, and the fact that Sidman's original subjects had moderate or severe MR, the literature on the MR population has grown slowly relative to that on normally capable individuals. Yet the study of individuals with MR has the potential to make important contributions. Because of their verbal limitations, their performances on tests for equivalence are potentially of both theoretical and practical importance.

Theoretically, studies of individuals with severe verbal limitations bear on a matter of fundamental importance to this research area. Although it is generally agreed that there is a close relation between equivalence relations and language, there is some disagreement as to the nature of that relation. One prominent view is that subjects' verbal behavior plays a necessary role in laboratory demonstrations of equivalence relations. Horne and Lowe (1996) proposed that performance on equivalence tests is made possible by naming. They carefully define the naming relation as a bidirectional one, involving both production of the name and responding to the referent upon hearing its name. They also propose two likely mechanisms by which stimulus naming promotes accurate equivalence test performance. One mechanism involves the subject emitting (often covertly) a common name to each stimulus in a stimulus class. Another possibility is that each stimulus in an eventual class controls a different name. Here, the proposed mechanism involves the development of intraverbal relations among stimulus class members. That is, emitting the name of one class member comes to control the emission of others, which in turn controls comparison selection.

In contrast, Sidman (1994) suggested that equivalence is a primary behavioral process not reducible to any other. To Sidman, equiv-

alence performances are a critical part of the foundation for language, not vice versa. Thus the demonstration of equivalence relations in individuals who do not name would be consistent with Sidman's view (although it would not prove the view). In contrast, the demonstration of equivalence relations by nonnaming subjects would provide evidence against the view that the naming repertoire is necessary to the laboratory equivalence performances.

There are several populations in whom the necessity-of-naming question should be investigated. These include preverbal normally developing children, individuals with MR and language limitations, and nonhuman animals. Demonstrations of equivalence in animals would provide the most convincing evidence that equivalence performances are not language driven. To date, clear and strong demonstrations have been limited, however. For animals, negative outcomes may be species related, although not necessarily an absolute limitation. On the one hand, the quest for reliable emergent performances in animals may involve, for example, exploring the use of more ethologically valid stimuli and other refinements in training procedures (see Sidman, 1994, pp. 163–164). On the other hand, some authors have questioned whether the same processes would be involved when, and if, there are reliable demonstrations of equivalence in nonhumans (Horne & Lowe, 1996).

Although arguably the most interesting, pre-naming children are notoriously difficult to study, and they are undergoing rapid developmental change. Moreover, it is unethical to deprive infants of primary reinforcers, and the powerful generalized conditioned reinforcers that often can be identified for adolescents and adults are not yet established in infants. Another human population in which extreme verbal limitations can be found consists of individuals with MR, who sometimes reach adolescence or adulthood with little-to-no verbal repertoire. In general, a greater variety of conditioned reinforcers have been established, and laboratory performance can usually be sustained for longer within-session and across-session periods. Moreover, adult subjects are unlikely to be undergoing developmental changes. On the one hand, however, one could argue that negative outcomes

from such individuals could be due to more than the absence of naming. On the other hand, positive outcomes also pose interpretation difficulties. One could argue that exposure to a verbal environment has produced naming abilities to which conventional assessments were not sensitive.

Experimentation and interpretation difficulties abound. There appears to be no perfect subject for this work. These considerations suggest that it is important to study all three groups—each can make a unique contribution. They are not substitutable for one another, but rather complementary. One of the goals for this review is to critique the existing literature on one of the aforementioned populations, individuals with MR and language limitations, in terms of whether this literature can be brought to bear on the necessity-of-naming question.

The three populations we have been discussing are of theoretical interest for another reason. They may provide the opportunity to directly study the development of emergent performances. The conditions sufficient for the initial development of these skills can best be studied in subjects who do not initially demonstrate them, and the study of the aforementioned populations may uncover such individuals. For example, Schusterman and Kastak (1993) did not initially show strong evidence of symmetry in their sea lion subject. After directly training symmetric relations, however, symmetry was shown in subsequent tests with new stimuli. This outcome is supportive of relational frame theory, which posits that a history of reinforcement of bidirectional relations is necessary for the demonstration of symmetry (Hayes, Barnes-Holmes, & Roche, 2001). Thus the second set of goals for this review is to determine (a) whether the literature has uncovered individuals who have failed to demonstrate equivalence, and (b) whether there have been attempts to analyze failures, or to produce the emergent performances that define equivalence relations in individuals who have not shown them.

Perhaps the most important reason for studying equivalence in this population is an immensely practical one. Tests for equivalence may represent a valuable tool in the study of language development (Goldstein, 1993). For example, they may provide a

means of assessing symbol comprehension in individuals with MR (Brady & McLean, 2000). The tests may be especially valuable for individuals who use alternative communication systems such as pointing to a picture or symbol on a communication board. Developing methods for assessing comprehension is a recently stated research priority of the National Institute for Deafness and Other Communication Disorders. It is worth noting that tests with similar rationale and structure have been used in language research with primates for some time (Savage-Rumbaugh, Rumbaugh, Smith, & Lawson, 1980). The equivalence relations between symbols and familiar objects might be maximally relevant to the applied study of symbol comprehension.

These considerations suggest that it is timely to review the literature on equivalence relations in individuals with MR who have language deficits. This review is structured around three features of the literature. First, we look at the characteristics of the subjects and how well characteristics have been described. Importantly, we look at how well the subjects' language abilities have been described. Language skills can vary widely even within a given level of MR (McLean, Brady, & McLean, 1996). As we have discussed, adequate information on language skills could make the data in this literature relevant to important theoretical and practical issues. There are other features of subject description that could bear on the interpretation of study outcomes. Have research reports indicated how the subjects were selected and whether or not they had previous experience with matching-to-sample procedures? Do the papers report level of MR, reported IQ, other diagnoses (such as autism or Down syndrome), and whether or not subjects were concurrently taking psychoactive medication?

Second, we examine the nature of the procedures used. We summarize the procedures in terms of whether or not they build on performances already in the subject's repertoire, whether all of the experimentally relevant relations were learned within the context of the experiment, and whether relations learned within the experiment included stimuli with which the subject was familiar. Third, we summarize test outcomes.

Although our primary interest here is in individuals with MR and limited language, we

found that the literature is small and descriptions of language skills are often incomplete. Thus we also include all reports of subjects with MR below the mild level because many, if not most, of such individuals exhibit serious language limitations. We found 20 studies that included at least 1 subject with moderate, severe, or profound MR. Some of these studies also included subjects with mild MR; we have not included data from these subjects in our summary. We also have not included a study by de Rose, McIlvane, Dube, and Stoddard (1988) because it was a follow-up of a previous study that involved similar methods and the same subjects (de Rose, McIlvane, Dube, Galpin, & Stoddard, 1988). For 3 other subjects who appeared in more than one study, however, both sets of data are included. These cases involved the demonstration of equivalence under two different procedures; this will be discussed further in a later section. Excluding subjects with mild MR, the 20 studies included 55 individual subjects. Because 3 of these individuals participated in more than one study, there were 58 instances of equivalence testing. We focus on individual performance, however, so even though all 58 cases are reported in the tables below, we discuss the outcomes of the review in terms of the 55 subjects.

The raw data from the articles reviewed are presented in tables. For manageability as well as expository purposes, the primary table is broken into three parts on the basis of differences in method (Table 1 a through c). We will describe these differences more fully later, when they are relevant to the discussion of training and testing methods. In what follows, we present what we found in the literature. The first three sections are largely a summary. We summarize the characteristics of the subjects, we summarize and categorize the procedures used, and finally we summarize the outcomes. In the last section, we discuss the state of the literature as well as some possible future research directions.

## SUBJECT CHARACTERISTICS

### *Demographics*

For each “qualifying” subject, Table 1 presents sex and age (in parentheses after subject code), level of MR, IQ or mental age (MA) and how it was assessed, other diagno-

ses, and language skills. Note that we report a level of MR for all subjects even though it was provided in the articles for only about half of the subjects. When the level of MR was not provided but a standardized test score was given, we assigned a level of MR based on that score using the scale provided by the American Psychiatric Association in the Diagnostic and Statistical Manual of Mental Disorders (1994). When the level of MR was provided, we used that level even if it conflicted with reported IQ or MA.

Table 2 summarizes the distribution of basic demographic characteristics (sex, age, and level of MR) for the 55 subjects. Perhaps the most striking aspect of these data is that there are few subjects in each of the demographic categories. The number of females (13) is especially low. Also noteworthy is the near absence of subjects with profound MR.

Returning to the information provided in Table 1, specific diagnoses were reported for only 21 subjects. Because nearly half of these 21 had Down syndrome, individuals with diagnoses other than Down syndrome may be underrepresented in the literature. No other specific physical disorder, including seizure disorder, was reported for more than 2 of the subjects. Also lacking is information on whether or not subjects were taking psychoactive medications. Only McDonagh, McIlvane, and Stoddard (1984) reported this information. Given that the prevalence of psychoactive medication use among individuals with MR ranges from 26 to 50% (Aman & Singh, 1988), it seems likely that more of these subjects were taking such medication. It is presently unknown whether these subject attributes—diagnosis, presence or absence of seizure disorder, and whether or not medication is used—affect performance on equivalence procedures. As the literature builds, however, presenting this information in all published reports may allow meta-analyses on the basis of these characteristics.

### *Preexisting Skills*

Thirty of the 42 subjects for whom information on preexisting skills was reported had a prior history with generally relevant experimental tasks (e.g., exposure to apparatus, identity matching to sample [IDMTS]). More directly relevant, and shown on the table, is the subjects' history of arbitrary matching to

sample (AMTS). Out of the 35 subjects for whom the presence or absence of AMTS experience was reported, 22 had demonstrated AMTS performance prior to participation in the studies reported here (3 subjects had demonstrated AMTS performance in previous studies: Subject DL in de Rose, McIlvane, Dube, & Stoddard, 1988; and Subjects PNO and JDB in Dube & McIlvane, 1995).

In general, the reports reviewed here provided little information as to whether preexperimental history or particular preexisting skills were factors in subject selection. Although subjects who do not learn AMTS cannot be given the classic tests for equivalence, few studies have reported explicit screening procedures for AMTS skills, or whether or not prospective subjects who did not learn AMTS were dropped. What can be gleaned from some of the studies is the number of trials required for the subject to learn the first conditional discrimination and the methods used to teach it. Most of the studies were completed before the relatively recent development of teaching procedures specifically designed for conditional discrimination (K. J. Saunders & Spradlin, 1990, 1993; Zygmunt, Lazar, Dube, & McIlvane, 1992). Appendixes A, B, and C (7th column) show that AMTS performances were most often taught either with trial-and-error procedures or with relatively crude instructional programming. For example, subjects learned with programming originally designed to teach simple discriminations, such as the delayed prompt procedure or fading in comparison stimuli. Table 1 a through c (column 9) shows that, with a few exceptions, these procedures were usually effective in 300 trials or fewer (although a number of studies did not report this information). Thus it may be that most studies have not included difficult-to-teach subjects.

#### *Communication Skills*

As we have noted, documentation of subjects' language ability could be an extremely valuable addition to studies of equivalence relations. Further, it would be helpful if part of that documentation came from standard assessment tools that involve direct measures of specific behavior (i.e., not based on observer ratings), so that comparable information would be provided across studies. In the existing literature, however, only two studies

have given formal tests of expressive language (Brady & McLean, 2000; Carr, Wilkinson, Blackman, & McIlvane, 2000), so we have not included a column for this on the table. A formal test of receptive skills, however, was common. The Peabody Picture Vocabulary Test (PPVT) was given to 39 (71%) of the subjects. The PPVT requires pointing to a picture (from among four choices) that corresponds to a spoken word.

In addition to formal tests, informal reports of communication skills were provided for 35 subjects. Sources of this information were either observations made by experimenters or reports by teachers or caregivers. Because these were idiosyncratic in form, we transcribed relevant information from the published study directly onto the table. Of these 35 subjects, 13 emitted complete sentences, 5 emitted single words or simple phrases, 10 communicated with nonvocal means (i.e., signs, gestures, or augmentative devices), 4 emitted unintelligible vocalizations, and 1 was mute.

#### PROCEDURES

In this section, we divide the procedures into categories based on the nature of the stimuli used and the likelihood that subjects had previous experience with the stimuli. Subject and outcome information for each category is presented in a separate section (a, b, and c) of Table 1. Additional information about the procedure for each study is presented in the three Appendixes: A, B, and C; that is, for each section of Table 1 there is a corresponding appendix.

There were three major categories. One included studies in which all of the stimuli were presumably new to the subjects; for example, Greek letters (Table 1a). The second included studies in which stimulus classes were established using a differential-outcomes procedure (Table 1b). Such studies are formally distinctive in that the reinforcer becomes a stimulus class member. In the third procedural category, already established conditional relations involving familiar stimuli were among the prerequisite relations for the target emergent behavior (Table 1c). Studies incorporating a previously established conditional relation are distinctive in that they may involve the addition of new stimuli to existing stimulus classes. It has been suggested that

Table 1a  
Studies involving novel relations with novel stimuli.

Study and subject	Mental retardation	IQ/mental age	Other diagnoses	Arbitrary matching to sample experience
Carr et al. (2000) Exp. 2 BN (M, 13)	S		Autism	
HF (M, 14)	S		Autism	
de Rose, McIlvane, Dube, Galpin, and Stoddard (1988) Exp. 3 MS (M, 31)	M	WAIS 46		n
DL <sup>a</sup> (M, 25)	M			y
Devany et al. (1986) Carl (M, 3-3)	M <sup>b</sup>	SB/Bayley 20 mo.		
Craig (M, 4-4)	S <sup>b</sup>	SB/Bayley 18 mo.		
Debbie (F, 2-7)	S <sup>b</sup>	SB/Bayley 14 mo.		
Sidman et al. (1974) PA (M, 18)	S		Down	y
Sidman, Kirk, and Willson-Morris (1985) PM (M, 21)	S <sup>b</sup>		Down	
Sidman et al. (1986) JL (M, 19-25)	S <sup>b</sup>		Down	
PA (M, 19-25)	S <sup>b</sup>		Down	
AA (M, 19-25)	S <sup>b</sup>		Hydrocephalus	
Spradlin and Saunders (1986) Exp. 4 BJ (M, 13)	M <sup>d</sup>	40		
VanBiervliet (1977) 1 (M, 21)	M <sup>d</sup>	48 <sup>g</sup>		
2 (M, 14)	M <sup>d</sup>	46 <sup>g</sup>		
3 (M, 11)	S <sup>d</sup>	35 <sup>g</sup>		
4 (M, 13)	P <sup>d</sup>	16 <sup>g</sup>		
5 (M, 13)	S <sup>d</sup>	27 <sup>g</sup>		
6 (M, 17)	S <sup>d</sup>	33 <sup>g</sup>		

Table 1a  
(Extended)

Language skills	Peabody Picture Vocabulary Test	Name	Number of trials	Test results			
				S	T	E	%N
Devonshire age equivalence = 2-0; no speaking repertoire; imitates animal sounds; manual signing; severely limited listening skills			70	+		(+)	
Reynell age equivalence = 3-1; speech mainly echoic or nonsensical; sometimes completes familiar sentences; functional speech was a few stock mands			620	X		X	
Communicative speech with phrases and simple sentences			128	+		+	
Single words, simple phrases	4.7		128	+		+	
Complete, brief sentences when prompted; spontaneous asking for items, commenting on classroom events			90			X	
Vowel sounds; no functional speech; no use of signs or picture boards			200			X	
Echolalic without comprehension; no functional speech; no use of signs or picture boards			360			X	
Considerable oral-naming experience		n		+	I		60
	4-1					+/I <sup>c</sup>	
	4			+		+/I <sup>c</sup>	86/0 <sup>c</sup>
	3-1			I		(+)/I <sup>c</sup>	33/0 <sup>c</sup>
	4-5			+		+/(+) <sup>c</sup>	100/0 <sup>c</sup>
				(+)		+	
Some receptive and productive language; names variety of objects; answers simple questions; initiates conversation				+		+	87
(same)				+		I	87
(same)				+		+	100
(same)				+		+	100
(same)				+		+	100
(same)				+		+	73

Note. Explanations and footnotes appear at the end of Table 1c.

Table 1b  
 Studies training novel relations with different outcomes.

Study and subject	Mental retardation	IQ/mental age	Other diagnoses	Arbitrary matching to sample experience
Dube et al. (1987) Exps. 1 and 2 DL <sup>a</sup> (M, 23)	M			y
MC (F, 43)	S		Down	y
Dube et al. (1989) PNO <sup>e</sup> (M, 20)	M		Seizure	
JDB <sup>e</sup> (M, 14)	M		Seizure & BD	
Dube and McIlvane (1995) Exp. 1 PNO <sup>e</sup> (M, 20)	M		Seizure	y
JDB <sup>e</sup> (M, 14)	M		Seizure & BD	y
JYT (M, 18)	S			y
RDP (M, 16)	M		BD	n
Dube and McIlvane (1995) Exps. 1 and 2 DGA (M, 16)	M <sup>b</sup>			y
DBH (M, 21)	M		Autism	n
LSM (M, 19)	S <sup>b</sup>		Autism	n
ADY (F, 13)	S		BD	n

the addition of stimuli to already existing classes may occur more readily than the development of classes “from scratch” (Dixon & Spradlin, 1976; R. R. Saunders, Wachter, & Spradlin, 1988).

There is a large literature on equivalence relations beyond the studies reviewed here, and procedures usually fit within our first category. That is, all prerequisite conditional relations are established in the laboratory. The established relations are between stimuli that are presumably unfamiliar to the subjects (e.g., Greek letters). These procedures are used to minimize effects of preexperimental

histories. In addition, in most studies involving normally capable subjects, all of the stimuli are visual (probably for convenience). In much of this literature, subjects are verbally sophisticated. Often the subjects are college students. Studies involving individuals with MR in which novel relations were trained with unfamiliar stimuli are shown in Table 1a. Note that one of these studies provided a within-subject comparison of the development of emergent performances with exclusively visual stimuli versus classes involving one auditory member (Sidman, Willson-Morris, & Kirk, 1986).

Table 1b  
(Extended)

Language skills	Peabody Picture Vocabulary Test	Name	Number of trials	Test results			
				S	T	E	%N
Answers simple questions about daily activities	4-7		18	+	+	+	100
Follows simple spoken instructions; speech mostly unintelligible	2-7		28	+	+	+	100
	4-1		44	+	+	+	
	3-8		34	+	I	+	
Vineland age-equivalent communication: 3.4	4.1		na			+	
Vineland age-equivalent communication: 5.5 Preschool Language Scale overall age score: 4.2	3.8		na			(+)	
	3.8		na			+	
	3.1		na			+	
	7.5		na			X/+ <sup>f</sup>	
Test for Auditory Comprehension of Language age-equivalent: 3.67	3.4		na/540x <sup>f</sup>				
	3.1		na/360x <sup>f</sup>				
Mute; Vineland age- equivalent communication: 2.4	3.6		na	+		X/X <sup>f</sup>	

*Note.* Explanations and footnotes appear at the end of Table 1c.

Sidman's seminal demonstrations of equivalence relations in subjects with MR differ from those just discussed in that they built on at least one conditional relation that existed preexperimentally. Studies using such procedures are shown in Table 1c. The relations that existed preexperimentally are shown in italics in the table in Appendix C. For example, in Sidman's early studies, one of the conditional discriminations that served as a prerequisite for equivalence tests involved spoken-word samples and picture comparisons. The 5 subjects in these studies demonstrated 40-75% accuracy (in an eight-choice

task) on these relations prior to training (Sidman, 1971; Sidman & Cresson, 1973; Sidman, Cresson, & Willson-Morris, 1974). In other studies, larger existing equivalence classes were expanded (Gast, VanBiervliet, & Spradlin, 1979; Maydak, Stromer, Mackay, & Stoddard, 1995; McDonagh et al., 1984). For example, before beginning the study by McDonagh et al., the subject was able to match coins to dictated and printed values, and name and/or state the value of coins and printed values. The target performances involved combinations of coins; for instance, matching two nickels to a dime.

Table 1c  
Studies involving preexisting relations.

Study and subject	Mental retardation	IQ/mental age	Other diagnoses	Arbitrary matching to sample experience
Brady and McLean (2000) TC (F, 33)	S			n
FR (M, 45)	S			n
VB (M, 26)	S			n
WB (M, 52)	S			n
RG (F, 30)	S			n
LT (F, 31)	S			n
BJ (M, 31)	S			n
BD (M, 28)	S			n
Carr et al. (2000) Exp. 1 DJB (M, 19)	S			y
JRV (F, 15)	S			y
IVB (M, 21)	S			
Gast et al. (1979) 5 (NR, 11-1)	M <sup>d</sup>	WISC 50		y
Mackay (1985) A (M, teen <sup>1</sup> )	S		Down	y
B (M, teen <sup>1</sup> )	S		Down	y
C (M, teen <sup>1</sup> )	S		Microcephaly	y
Mackay and Ratti (1990) A (F, 72)	S	Leiter		y
B (F, 60)	S	Leiter		y
C (F, 49)	S	Leiter		y
Maydak et al. (1995) Brad (M, 30)	M <sup>b</sup>			
Mary (F, 49)	S <sup>b</sup>			
McDonagh et al. (1984) (F, 28)	M			y

Table 1c  
(Extended)

Language skills	Peabody Picture Vocabulary Test	Name	Number of trials	Test results			%N
				S	T	E	
Produced grammatically correct sentences, some with embedded clauses	3-7	y	255	+	+		100
(same)	2-8	y	1377	I	I		87
(same)	2-8	y	270	+	+		100
(same)	4-11	y	750	+	X		70
No speech or other symbolic communication; communicated with nonsymbolic gestures	4-2	n	210	+	I		
(same)	4-2	n	805	+	X		
(same)	2-8	n	2140	+	+		
(same)	2-8	n	4150x				
EOWPVT-R <sup>h</sup> = 2 years 5 months; limited signing; picture communication board; mostly mands; no significant oral naming skills	2-3		~4800	+	+	+	
EOWPVT-R <sup>h</sup> <2 years limited signing; mostly mands; no significant oral naming skills	2-3		~240	+	+	+	
EOWPVT-R <sup>h</sup> <2 years; picture communication board; mostly mands; no significant oral naming skill	2-1		~960	(+)	+	+	
Counts numerals		y				(+)	
No writing		y	na	I	I		100
(same)		y	na	+	I		100
(same)		y	na	I	I		78
	3-3	y		+		+	100
	5-3	y		+		+	100
	2-4	y		+		+	100
Reading level grade 2.1	7-9			(+)	(+)	(+)	100
Nonreader	3-6			+	+	+	93
Good receptive and expressive oral language skills; no reading or writing		y				+	100

Table 1c  
(Continued)

Study and subject	Mental retardation	IQ/mental age	Other diagnoses	Arbitrary matching to sample experience
Sidman (1971) (M, 17)	M <sup>j</sup>	WAIS 49 <sup>k</sup>	Microcephaly	y
Sidman and Cresson (1973) BA (M, 19)	S	4-2 <sup>l</sup>	Down	
JW (M, 18)	S	4-0 <sup>l</sup>	Down	
Sidman et al. (1974) JC (M, 14)	S	4-11 <sup>l</sup>	Down	y
Stromer and Mackay (1993) TR (F, 14)	M <sup>d</sup>	WISC-R 47		y
KB (M, 20)	S <sup>b</sup>			y

*Note.* A blank cell indicates that information was not provided or tests were not given. Subject identifier is followed by gender and age in parentheses. In the *Mental Retardation* column, *M* = Moderate, *S* = Severe, and *P* = Profound. In the *IQ/Mental Age* column, *WAIS* = Weschler Adult Intelligence Scale; *SB* = Stanford-Binet IQ Scale; and *WISC* = Weschler Intelligence Scale for Children. In the *Other Diagnoses* column, *Down* refers to Down syndrome; *BD* refers to Behavior Disorder. The *Arbitrary Matching to Sample (AMTS) Experience* column indicates whether subjects had experience with AMTS procedures prior to participation (not including pretraining). Information in the *Language Skills* column is a slightly abbreviated transcript of language information presented in the article. The *Name* column indicates whether the subject had demonstrated naming of experimental stimuli (approximately 50% or more) prior to any training. The number in the *Number of Trials* column indicates the number of trials to mastery criterion on the first arbitrary conditional discrimination, if given, except for Brady and McLean (2000), which is the total number of trials for six relations (the number is followed by *X* if the subject never met criterion). For *Test Results*, *S* = Symmetry, *T* = Transitivity, *E* = Equivalence, and *%N* indicates the percentage of naming responses consistent with the auditory member of the class; accuracy on the last test is reported. For naming, scores on the last test for emergent naming are reported. For symbols in *Test Results*: + designates accuracy of 90% or above, *I* designates intermediate accuracy (the text indicates how this was determined), and *X* designates below intermediate accuracy. If symbols or numbers are in parentheses, there was gradual emergence of terminal performance.

<sup>a</sup> Participated in Dube et al. (1987) and de Rose, McIlvane, Dube, Galpin, and Stoddard (1988).

<sup>b</sup> Level interpreted from Peabody Picture Vocabulary Test age equivalent or Stanford-Binet score.

<sup>c</sup> Auditory-visual/visual-visual tests.

<sup>d</sup> Level interpreted from reported IQ.

<sup>e</sup> Participated in Dube et al. (1989) and Dube and McIlvane (1995).

<sup>f</sup> Exp. 1/Exp. 2 results. With relations involving food, ADY showed equivalence only when food was presented as the sample.

<sup>g</sup> Based on Peabody Picture Vocabulary Test.

<sup>h</sup> Gardner Expressive One-Word Vocabulary Test-Revised.

<sup>i</sup> No age reported.

<sup>j</sup> Labeled "moderate" in Sidman and Cresson (1973), but labeled "severe" in Sidman (1971).

<sup>k</sup> As reported in Sidman and Cresson (1973).

<sup>l</sup> Stanford Binet: No numerical IQ given.

Most studies that built on preexisting relations also included one auditory stimulus, and Table 1c shows that subjects often began the study naming the visual stimulus to which the auditory stimulus was related (indicated by a "y" in the "Name" column). The previous paragraph contained an example from

Sidman's early studies; in addition to selecting the picture upon hearing its name, subjects named many of the pictures prior to training. As noted previously, there is some evidence that emergent performances are more likely when the stimulus class has an auditory member (Sidman et al., 1986; see

Table 1c  
(Continued, Extended)

Language skills	Peabody Picture Vocabulary Test	Name	Number of trials	Test results			%N
				S	T	E	
Good auditory comprehension and picture naming; little or no reading comprehension or oral reading; no writing		y				I	75
		y				I	90
		y				I	65
Considerable oral-naming experience	3-4	y	545	+	I		55
Reading beginning 3rd grade	5-11				+	+	89
Unintelligible speech	4-1				+	+	

also Green, 1990). It has also been theorized that emergent performances depend on naming (Horne & Lowe, 1996); therefore, experimental arrangements that make naming more likely may increase the likelihood of emergent performances.

Our notion of adding to existing classes may oversimplify cases in which classes that are built on preexisting relations involve more than three stimuli. For example, for the subjects in Experiment 1 of Carr et al. (2000), spoken-word-to-picture relations were entry skills, and two new conditional discriminations were taught; one with printed-letter

samples and the pictures as comparisons and one with arbitrary-stimuli samples and picture comparisons. The subjects had no preexisting relations involving either the printed letters or the arbitrary stimuli. Thus emergent relations between the printed letters and arbitrary stimuli did not directly include a stimulus that was in a preexisting relation (as would be the case if the class had only three stimuli), but they *were* based on the relation of these two stimuli to the same picture.

A third way that the choice of experimental stimuli can vary across studies might be seen as a hybrid of the previous two examples. In some studies, all of the prerequisite relations were taught in the study (i.e., there was no preexisting relation), but at least one stimulus in each prospective stimulus class was an everyday object or picture that presumably was familiar to the subject. That is, relations between familiar objects and novel symbols were taught. These studies are shown in Table 1b. It turns out that the only examples of this category involve differential outcome procedures. In these, the familiar stimuli were the reinforcers, and a different reinforcer was associated with each sample-comparison relation in a trained conditional discrimination. There were three studies which included a total of 10 different subjects (Dube & Mc-

Table 2

Number of subjects according to level of mental retardation, age, and gender.

Age	Gender	Level of mental retardation		
		Moderate	Severe	Profound
Adult ( $\geq 18$ yrs)	M	6	17	0
	F	1	8	0
	Total	7	25	0
Child ( $< 18$ yrs)	M	7	10	1
	F	1	3	0
	Total	9 <sup>a</sup>	13	1
Grand total		16 <sup>a</sup>	38	1

<sup>a</sup> Totals include one moderate-level child missing from the tallies because no gender was reported.

Ilvane, 1995; Dube, McIlvane, Mackay, & Stoddard, 1987; Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989). A study by Dube and McIlvane provides an example. The 8 subjects learned two 2-choice identity-matching-to-sample problems. In each problem, correctly selecting one stimulus produced one type of reinforcer and correctly selecting the other stimulus produced a different type of reinforcer. Tests determined whether subjects matched the stimuli that produced the same outcome. For example, selection of A1 in the presence of A1 and B1 in the presence of B1 produced one consequence, R1 (e.g., chocolate cookie). Selection of A2 in the presence of A2 and B2 in the presence of B2 produced a different consequence, R2 (e.g., cheese cracker). In test trials, A1 and A2 were presented as a sample with B1 or B2 as comparison stimuli (i.e., tests were AMTS). Class-consistent responding was demonstrated if subjects matched A1 and B1 or A2 and B2.

It should be noted that 3 subjects who participated in the differential outcomes studies participated in more than one study reviewed here, and we show both data sets because of significant procedural differences. The 2 subjects in Dube et al. (1989) later participated in Dube and McIlvane (1995). Both studies involved differential outcomes, but in the former, the training initially included both identity and arbitrary relations and the arbitrary relations that were included provided a basis *other than differential outcomes* for some of the emergent performances. This experience may have increased the likelihood of emergent performances based on differential outcomes. In contrast, Study 1 of Dube and McIlvane, described above, provided the minimal training that is theoretically necessary for emergent relations based on differential outcomes. The third subject who appears twice participated in both a study involving differential outcomes (Dube et al., 1987) and a study that did not involve differential outcomes (de Rose, McIlvane, Dube, Galpin, & Stoddard, 1988).

#### EQUIVALENCE TEST OUTCOMES

The last four columns of Table 1a through c show the results of tests for symmetry, transitivity, equivalence (a combined test for symmetry and transitivity), and emergent naming (if tested). If, after initial positive testing, ad-

ditional conditional relations were taught and additional tests given, we have not included this information unless higher accuracy was shown. The symbols in the symmetry, transitivity, and equivalence columns represent terminal performance on the first set of tests (i.e., prior to any further training): + designates accuracy of 90% or above, *I* designates intermediate accuracy, and *X* designates failure to demonstrate emergent relations. The accuracy levels that led to *I* and *X* designations depended on the number of comparison stimuli (because the accuracy level that represents chance performance depends on the number of comparisons). The formula for calculating the lower limit of *I* performances was 100 plus chance divided by 2. For example, chance on a three-choice task is 33.33%, so *I* performances were those between 66.67% and 90% (100 plus 33.33 divided by 2 = 66.67). We defined *X* performance as that which fell below the *I* lower limit (e.g., in a three-choice task, *X* performance would be <66.67%).

In the literature reviewed here, 34 of the 55 subjects showed accuracy of above 90% across all equivalence tests and 16 showed intermediate accuracy on at least one test type (i.e., symmetry, transitivity, or equivalence). Nine subjects did not show at least intermediate accuracy in one or more equivalence tests. In some cases, subjects did not initially demonstrate the emergent performances of interest, but additional manipulations resulted in positive equivalence test outcomes (for example, Subjects DGA and ADY in Dube & McIlvane, 1995). Moreover, in some cases accuracy increased across tests. In both of these cases, we based our tally on terminal accuracy. Given the small number of subjects thus far involved, it is not possible to draw firm conclusions as to how subject or procedural variables might affect equivalence test outcomes. The review shows, however, a substantial number of subjects who did not show highly accurate performance. These data, and speculation as to the reason for them, may provide a basis for further analysis of the development of emergent stimulus control.

The negative cases are of particular interest. Three of the negative cases were associated with failure to establish arbitrary matching. One of these subjects never received tests for emergent performances (Subject BD in

Brady & McLean, 2000). For two other negative cases, initial training sought emergent relations between stimuli whose selection produced a common outcome (i.e., a differential outcomes procedure; Dube & McIlvane, 1995). The training prior to these tests involved identity rather than arbitrary matching, but the tests involved arbitrary matching. With these procedures, 4 of 8 subjects did not initially show emergent arbitrary matching based on differential-outcomes procedures. In a follow-up study that used the more usual arbitrary matching procedures, 2 of these subjects did not acquire arbitrary matching and are included in our total for negative outcomes (the other 2 learned arbitrary matching and showed positive test outcomes). It is possible that the initial test failures by 4 subjects, and subsequent failure to teach arbitrary matching to 2 of these, was due to the subjects' experimental history of reinforcement for identity matching. Stimulus control by physical identity may have competed with the demonstration of equivalence relations and with follow-up attempts to train arbitrary matching.

The other six failures, of the nine found in this review, did not demonstrate equivalence despite learning the arbitrary matching baseline. These six failures included subjects in three studies: Devany, Hayes, and Nelson (1986); Carr et al. (2000); and Brady and McLean (2000). Devany et al. showed negative outcomes on an equivalence test for three preschoolers with MR, two of whom were judged to be nonverbal. Aspects of Devany et al.'s methods, however, have led some to question whether these negative outcomes are convincing (McIlvane & Dube, 1996; K. J. Saunders & Spradlin, 1996). The accuracy criterion for beginning testing was not stringent, and the testing duration was brief relative to the number of sessions required for baseline acquisition. Thus the subjects had far more experience with responding incorrectly than correctly. Further, unlike almost all other studies, they discontinued exposure to the baseline trials once testing began. Thus it is unknown whether the prerequisite relations, which were necessary for positive test outcomes, were maintained at high accuracy. A more convincing negative outcome would have shown a discrepancy between highly accurate baseline performance and inaccurate

probe trial performance. (See Carr et al. for an excellent discussion of the status of Devany et al. in the literature.)

Two subjects in Brady and McLean (2000) showed symmetry, but not transitivity. The latter study apparently discontinued exposure to the baseline trials once testing began, and presented only one 30-trial test each for symmetry and transitivity. Thus these negative outcomes are also inconclusive.

Finally, 1 subject in Carr et al. (2000) passed neither a test for symmetry nor a test for equivalence (a combined test), despite showing high accuracy on the baseline relations throughout testing. Follow-up testing showed that this subject had intermediate-to-chance accuracy on tests for sample-S+ and sample-S- control. In sample-S+ control tests, the incorrect comparison is replaced with a novel form. High accuracy indicates a strong relation between the sample and the correct comparison (S+); responding does not depend on control by a specific incorrect comparison. In sample-S- tests, the correct comparison is replaced with a novel form. High accuracy shows control by exclusion of the incorrect comparison (S-) in the presence of that sample. Carr et al., note that the 3 other subjects had previously shown high accuracy on such tests (as part of another study), and all 3 showed highly accurate sample-S+ and S- control. Thus these tests show promise as a tool with which to analyze the prerequisite skills for positive test outcomes. McIlvane, Serna, Dube, and Stromer (2000) have proposed that negative outcomes on equivalence tests may often be due to a failure to establish the experimenter-intended baseline relations, and this example appears to be a case in point. It is important to emphasize that this failure can occur even though accuracy is 100%. O'Donnell and Saunders (1998) reported an example. The subject showed near perfect accuracy under a two-choice, visual-visual MTS procedure. Tests for sample-S+ and sample-S- control showed that he had learned only one of the two potential sample-S+ relations. When the second sample was presented, his selection reflected exclusion of the comparison that was correct in the presence of the first.

In summary, five of the negative outcomes involved extremely brief testing periods and discontinuation of the prerequisite baseline

trials once testing began, making the outcomes inconclusive. Three of the negative outcomes involved failures to establish arbitrary matching. Finally, additional stimulus control analysis of one negative outcome showed that strong sample-S+ and sample-S- relations had not been established.

In addition to the failures, it is important to note that 16 subjects showed intermediate accuracy (defined as accuracy below 90% but at least 1.5 times chance levels) on at least one test. Thus test performance involved a mixture of stimulus control relations, including relations not indicative of equivalence. Five of these, for example, involved subjects in Sidman's earliest studies, who showed intermediate accuracy in their final equivalence test (Sidman, 1971; Sidman & Cresson, 1973; Sidman et al., 1974). There were 20 different stimulus classes in these studies, and, as with some of the failure outcomes, only one test was presented after the prerequisites for all were taught. As in the failures discussed above, additional testing combined with exposure to the baseline might have resulted in the expected emergent performances. Such "gradual emergence" has been reported in studies that present tests multiple times (e.g., Spradlin, Cotter, & Baxley, 1973). Another feature of these particular studies that may have directly contributed to the "error" stimulus control relations is that the 20 baseline relations were taught gradually, whereas all potential emergent relations were tested periodically. Thus some potential emergent relations were tested both before and after their prerequisites were taught. It is possible that testing before prerequisites were taught established stimulus control that remained when the same tests were presented after the relevant training, competing with class-consistent responding. These papers do not describe the stimulus control shown on error trials, however, so this remains conjecture, albeit testable.

#### DISCUSSION

In 30 years of research on equivalence relations, there have been a number of important demonstrations. Many of these demonstrations have involved subjects with MR. The study of such individuals has the potential to make contributions of both a theoretical and practical nature. This review shows, however,

that in some ways the work has just begun. Despite the fact that the seminal studies in the area involved subjects with MR (i.e., Sidman, 1971; Spradlin et al., 1973), to date published reports over these 30 years have involved only 55 different subjects with moderate, severe, or profound MR. Because the literature is small and involves a range of subjects and procedures, the goal of this review is to provide a means of beginning to systematize this literature, and to point out areas in which the literature could be strengthened. In addition, the literature is beginning to provide data that address the theoretical and practical issues raised in the introduction.

Perhaps the most important theoretical issue in this research area as a whole (i.e., not specific to the population covered here) is that of whether equivalence test performances are mediated by language. Horne and Lowe (1996) have argued that equivalence performances depend on a history of learning to name, with the naming brought to bear on equivalence tests. When studies build on existing conditional relations that include a stimulus that the subject can name (e.g., Sidman, 1971), it is difficult to rule out this sort of verbal mediation in individuals with naming skills. Studies that show apparently nonexistent or demonstrably inconsistent naming (e.g., Sidman et al., 1986) provide some evidence against naming hypotheses. Studies with nonverbal individuals (along with animal studies), however, rule out mediation by naming, and lend credence to Sidman's notion that these emergent performances do not depend on a language-learning history. To date, there are only three published studies that have explicitly examined equivalence performance in individuals with extremely limited verbal repertoires. One study reported failure on equivalence tests (Devany et al., 1986), one reported positive results in 4 of 5 subjects (Carr et al., 2000), and one reported positive results in 2 of 4 subjects (Brady & McLean, 2000). A few other studies included a subject whose verbal skills appear to have been limited, but the description of the subjects' verbal skills was lacking.

Data thus far suggest that equivalence performances do not depend on naming, but there are few relevant studies. In essence, the

only human studies that support this conclusion are Brady and McLean (2000) and Carr et al. (2000). (Importantly, these are supplemented by demonstrations of equivalence in 2 sea lions [Kastak, Schusterman, & Kastak, 2001; Schusterman & Kastak, 1993].) Moreover, the Brady and McLean study was conducted with tabletop procedures in which an experimenter and the subject sat across from each other. Experimenter cuing is a risk in this arrangement. Steps were taken to diminish this possibility; the experimenter was told to look at the subject's face while selection was occurring, and the first author periodically watched sessions through a one-way mirror. Although no cuing was observed, any doubt would be eliminated by removing the experimenter from the subject's view and/or having a mechanically transduced response.

The Carr et al. (2000) study used computerized procedures for 3 of the 4 subjects who showed positive outcomes. One proviso, however, is that details as to the method of reinforcer delivery are not mentioned. Many studies from this laboratory involve experimenter-delivered reinforcers and note that the experimenter sat behind and to the side of the subject to avoid cuing. Such a procedure would diminish the possibility of cuing relative to a face-to-face tabletop procedure.

These sea lion and human studies provide important evidence, but the question requires additional work. Given the extreme practical and theoretical importance of studies of individuals with MR and language limitations, why have there been so few studies? The primary reason may be that this is slow and technically demanding work (also noted in Carr et al., 2000). One major roadblock is that it is difficult to establish arbitrary matching performances in such individuals. Unfortunately, this difficulty has generally not been documented in the equivalence literature itself. It has, however, inspired a parallel body of work by laboratories involved in the study of stimulus control processes in individuals with MR (McIlvane, Dube, Kledaras, Iennaco, & Stoddard, 1990; K. J. Saunders & Spradlin, 1989, 1990, 1993; Zygmont et al., 1992). Difficulties establishing arbitrary matching have almost certainly impeded research on equivalence relations in individuals with MR.

These arbitrary matching difficulties also raise the issue of whether the subjects that

have been studied are representative (also noted by Dube & McIlvane, 1995). It is possible that implicit screening may have occurred. That is, subjects may have been selected on the basis of known arbitrary matching abilities, or excluded (and their exclusion not reported) when they did not readily learn arbitrary matching (because it did not matter in answering the particular question posed in the study). Implicit screening seems possible for two reasons. The first, noted above, is that recent reports have brought to light difficulties in establishing arbitrary matching in this population, yet these difficulties are not represented in the published literature on equivalence relations. In general, the subjects in published studies readily acquired AMTS with relatively crude teaching procedures.

Second, when studies have explicitly reported AMTS training failures, the failure rate is fairly high. In addition to the previously mentioned studies, some studies with higher functioning subjects have reported failures. In Spradlin and Saunders (1986), 5 of 12 individuals did not learn AMTS; the functioning level of these individuals was not reported, but the majority of the other subjects in the study had mild MR. In Dube and McIlvane (1995), 2 of 4 subjects who successfully completed Experiment 1, which did not require AMTS, did not learn the basic AMTS task in Experiment 2. These two studies, along with several observations in other subject populations, suggest that this may be the tip of the iceberg. For example, Stromer and Osborne (1982) discontinued 9 of 22 subjects who did not achieve 90% accuracy on a conditional discrimination within 150 trials. Although this was not a lengthy training exposure, these were higher functioning subjects than are the focus of this review. In addition, work with young, developmentally normal children has also shown acquisition difficulties (Augustson & Dougher, 1991; Pilgrim, Jackson, & Galizio, 2000). In the present group of studies, however, only three teaching failures were reported (two in Dube & McIlvane and one in Brady & McLean, 2000).

The subject representativeness question suggests at least two courses of action: increased study of difficult-to-teach subjects and better documentation of subject selection procedures and skill levels. It is rare in this

literature for subject selection procedures to be reported.

To increase the number of individuals with severe MR in whom emergent stimulus control can be studied, two relatively new procedures have been designed specifically to establish AMTS in this population: K. J. Saunders and Spradlin's (1990, 1993) component training procedure, and Zygmunt et al.'s (1992) sample-shaping procedure. These procedures have established AMTS with individuals for whom trial-and-error procedures, or relatively crude instructional programming, have failed. Thus this area of research is poised to be extended to a greater number of subjects with severe language limitations. This will allow greater assurance that subjects with a range of preexisting capabilities are represented.

Along with the increased study of language-limited subjects, we strongly recommend that future studies provide better documentation of subjects' language capabilities. Standardized, replicable procedures for describing language skills would be the ideal. Standardized descriptions would facilitate the comparison of results across studies as well as the accumulation of sufficient data on which to base judgments of generality. It is a healthy sign that the majority of the studies reviewed here used a common measure of receptive language—the PPVT. Because the PPVT is a single-word test, it would be desirable to add measures of phrase and/or sentence comprehension, using, for example, the Test for Auditory Comprehension of Language (Carrow-Woolfolk, 1985).

Descriptions of expressive language have been extremely limited. Only two studies, Brady and McLean (2000) and Carr et al. (2000), have included standardized tests of expressive language. Another study, Devany et al. (1986), categorized subjects in terms of whether or not they had functional speech or sign skills. Three observers, who made this judgment independently, were in agreement as to each subject's status. Nonetheless, there is no information on the conditions under which subjects were observed, so the observation procedures cannot be replicated. A test that is appropriate for this purpose is the Expressive One-Word Vocabulary Test (Gardner, 1990). This test presents a series of pictures that the subject is required to name. (It

would be appropriate to allow signing as a response mode with such tests.) Carr et al. used this test for some subjects.

Brady and McLean (2000) used an innovative way of characterizing the behavior of language-limited subjects. Their test used analogue situations to assess rudimentary manding (requesting) and tacting (calling another person's attention to an unusual occurrence in the immediate environment). For example, the subject was put into a situation in which a tool needed for a particular task was missing, thus evoking a request (mand). In the rudimentary tacting analogue, an unusual item or event was scripted, and a trained observer noted whether or not the individual called his or her attention to the item or event. For example, a plastic alligator was dangled above the observer's head and, importantly, out of the observer's field of view. It is important to note that responses conventionally classified as nonverbal would be classified here solely in terms of their apparent function. For example, the subject might pound on the table and point to the dangling alligator until the experimenter oriented toward it. In Brady and McLean, 2 of 4 subjects who did not speak or sign, but who showed such rudimentary tacting, demonstrated symmetry and transitivity.

In a survey of the language skills of 28 individuals with profound MR, Brady, McLean, McLean, and Johnston (1995) found only 1 subject who did not mand, but 10 subjects did not show rudimentary tacting. Given that the floor skill for the most commonly used standardized, direct measures of subjects' expressive language is naming, continued development of these testing procedures could fill an important gap in our characterization of the verbal abilities of individuals with severe and profound MR. It is interesting to consider whether individuals whose verbal repertoires are limited to mands would perform differently than those with rudimentary tacting repertoires on tests for emergent stimulus control. No study has yet addressed this question. Recent improvements in establishing arbitrary matching, along with improvements in documenting subjects' verbal repertoires, may increase the likelihood of thoroughly investigating this question.

In addition to bearing on theoretical issues, studies of verbally limited individuals

could have immensely practical implications. For example, equivalence tests could be used as tests for the comprehension of symbols that are learned as part of an augmentative communication system. Ironically, despite the prominent notion that the definition of equivalence relations captures a fundamental feature of symbol-referent relations, there have been few studies in which subjects were taught relations involving a previously unnamed object in their natural environment. Such work would be directly applicable to individuals whose language delays involve comprehension difficulties.

A study by Brady and McLean (2000) shows this potential. They presented a set of tests that incorporated thematic classes that had formed prior to the experiment. (We have not shown the outcome of these tests on the table; the subjects also received basic equivalence tests, the outcome of which are shown on the table.) Example thematic classes included shoe and sock, and spoon and bowl. The subjects learned to select a different symbol in the presence of each of six objects. Tests then determined whether the subjects also matched, for example, the symbol for the shoe and the symbol for the sock.

Practical matters would also be addressed by further analyses of failures to demonstrate positive test outcomes (see McIlvane et al., 2000). These analyses would inform the development of improved teaching procedures. For example, O'Donnell and Saunders (1998) demonstrated that a subject who was highly accurate on a two-choice AMTS problem had actually learned only one of the sample-choice relations and rejected that choice when the other sample was present. Because this exclusion strategy inadvertently resulted in selecting the correct choice, accuracy was high. It is noteworthy that all subjects in Carr et al. (2000) received similar tests to characterize the nature of the taught arbitrary matching performance. Of the 5 subjects, the one who showed deficits on these tests also did not show high accuracy on the equivalence tests. Matching-to-sample teaching procedures are already widely used in teaching individuals with MR and severe language limitations. Procedures that ensure that all sample-choice relations have been learned would enhance the viability of MTS and equivalence

procedures for teaching individuals with limited language.

While on the topic of teaching procedures, it is worth noting that all of the reported failures to demonstrate equivalence in subjects who have mastered the prerequisite relations have involved tabletop testing procedures. Moreover, most of these have involved a brief testing phase that did not integrate review of the prerequisite conditional relations (i.e., in Devany et al., 1986, and Brady & McLean, 2000). Either or both of these features may have played a role in the negative outcomes. Tabletop training and testing, for example, may increase the likelihood that responding is based on features other than those desired by the experimenter (see McIlvane et al., 2000), such as subtle experimenter cuing (K. J. Saunders & Williams, 1998).

Although this review has focused primarily on language abilities, it is worth noting that a number of other subject characteristics may affect equivalence-test performance. Of course, targeted studies would be required to determine effects of additional factors, but it would seem worthwhile (and require little effort) to include potentially important information in each published report. For example, there is evidence that psychoactive medications can affect discrimination learning (Williams & Saunders, 1997). Almost none of the published studies reported whether or not subjects were being treated with psychoactive medications, although a sizable proportion of individuals with MR are undergoing such treatment. Would psychoactive medications affect the likelihood of emergent stimulus control? It is interesting to consider the use of equivalence tests as dependent measures in studies of medication effects.

Equivalence tests conceivably might be sensitive to different diagnoses within developmental disabilities. Would equivalence test outcomes differentiate children with autism, for example, given the language difficulties associated with this disorder? In the literature covered by this review, only 4 subjects were documented to have autism. Another question related to diagnosis is that a large proportion of the subjects in these studies have Down syndrome. Would the findings of the literature as a whole differ with a broader range of diagnoses?

In summary, the study of equivalence relations in individuals with moderate, severe, and profound MR holds both practical and theoretical importance. The existing research suggests that positive equivalence test outcomes can be shown in individuals with extreme verbal limitations. Difficulties in teaching arbitrary matching have slowed research in this area. In the last decade or so, however, great progress has been made in applying analyses of arbitrary matching performance to develop reliable teaching procedures. Thus this research area is poised for further productive study of individuals with MR and language limitations.

## REFERENCES

- Aman, M. G., & Singh, N. N. (1988). Patterns of drug use, methodological considerations, measurement techniques, and future trends. In M. G. Aman & N. N. Singh (Eds.), *Psychopharmacology of the developmental disabilities* (pp. 1–28). New York: Springer-Verlag.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Augustson, K. G., & Dougher, M. J. (1991). Teaching conditional discrimination to young children: Some methodological successes and failures. *Experimental Analysis of Human Behavior Bulletin*, 9, 21–24.
- Brady, N. C., McLean, J. E., McLean, L. K., & Johnston, S. (1995). Initiation and repair of intentional communication acts by adults with severe to profound cognitive disabilities. *Journal of Speech and Hearing Research*, 38, 1334–1348.
- Brady, N. C., & McLean, L. K. (2000). Emergent symbolic relations in speakers and nonspeakers. *Research in Developmental Disabilities*, 21, 197–214.
- Carr, D., Wilkinson, K. M., Blackman, D., & McIlvane, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior*, 74, 101–114.
- Carrow-Woolfolk, E. (1985). *Test for Auditory Comprehension of Language: TACL-R Examiner's Manual* (Rev. ed.). Allen, TX: DLM Teaching Resources.
- de Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior*, 50, 1–20.
- de Rose, J. C., McIlvane, W. J., Dube, W. V., & Stoddard, L. T. (1988). Stimulus class formation and functional equivalence in moderately retarded individuals' conditional discrimination. *Behavioural Processes*, 17, 167–175.
- Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior*, 46, 243–257.
- Dixon, M. H., & Spradlin, J. E. (1976). Establishing stimulus equivalence among retarded adolescents. *Journal of Experimental Psychology*, 21, 144–164.
- Dube, W. V., & McIlvane, W. J. (1995). Stimulus-reinforcer relations and emergent matching to sample. *The Psychological Record*, 45, 591–612.
- Dube, W. V., McIlvane, W. J., Mackay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior*, 47, 159–175.
- Dube, W. V., McIlvane, W. J., Maguire, R. W., Mackay, H. A., & Stoddard, L. T. (1989). Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior*, 51, 65–76.
- Gardner, M. F. (1990). *Expressive one-word picture vocabulary test-revised*. Novato, CA: Academic Therapy Publications.
- Gast, D. L., VanBiervliet, A., & Spradlin, J. E. (1979). Teaching number-word equivalences: A study of transfer. *American Journal of Mental Deficiency*, 83, 524–527.
- Goldstein, H. (1993). Structuring environmental input to facilitate generalized language learning by children with mental retardation. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Research foundations for intervention* (Vol. 2, pp. 317–334). Baltimore, MD: Paul H. Brookes.
- Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation*, 95, 260–270.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. The Netherlands: Kluwer Academic Publishers.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 185–241.
- Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior*, 76, 131–158.
- Mackay, H. A. (1985). Stimulus equivalence in rudimentary reading and spelling. *Analysis and Intervention in Developmental Disabilities*, 5, 373–387.
- Mackay, H. A., & Ratti, C. A. (1990). Position-numeral equivalences and delayed position recognition span. *American Journal on Mental Retardation*, 95, 271–282.
- Maydak, M., Stromer, R., Mackay, H. A., & Stoddard, L. T. (1995). Stimulus classes in matching to sample and sequence production: The emergence of numeric relations. *Research in Developmental Disabilities*, 16, 179–204.
- McDonagh, E. C., McIlvane, W. J., & Stoddard, L. T. (1984). Teaching coin equivalences via matching to sample. *Applied Research in Mental Retardation*, 5, 177–197.
- McIlvane, W. J., & Dube, W. V. (1996). Naming as a facilitator of discrimination. *Journal of the Experimental Analysis of Behavior*, 65, 267–272.
- McIlvane, W. J., Dube, W. V., Kledaras, J. B., Iennaco, F. M., & Stoddard, L. T. (1990). Teaching relational discrimination to individuals with mental retardation: Some problems and possible solutions. *American Journal on Mental Retardation*, 95, 283–296.
- McIlvane, W. J., Serna, R. W., Dube, W. V., & Stromer, R. (2000). Stimulus control topography coherence and stimulus equivalence: Reconciling test outcomes with theory. In J. C. Leslie & D. Blackman (Eds.), *Experimental and applied analysis of human behavior* (pp. 85–110). Reno, NV: Context Press.

- McLean, L. K., Brady, N. C., & McLean, J. E. (1996). Reported communication abilities of individuals with severe mental retardation. *American Journal on Mental Retardation*, 100, 580–591.
- O'Donnell, J., & Saunders, K. J. (1998). An attempt to change inadvertently established sample-S control. *Experimental Analysis of Human Behavior Bulletin*, 16, 7–9.
- Pilgrim, C., Jackson, J., & Galizio, M. (2000). Acquisition of arbitrary conditional discriminations by young normally developing children. *Journal of the Experimental Analysis of Behavior*, 73, 177–193.
- Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: The effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior*, 52, 1–12.
- Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior*, 54, 239–250.
- Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, 60, 571–585.
- Saunders, K. J., & Spradlin, J. E. (1996). Naming and equivalence relations. *Journal of the Experimental Analysis of Behavior*, 56, 304–308.
- Saunders, K. J., & Williams, D. C. (1998). Stimulus control procedures. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 193–228). New York: Plenum.
- Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 49, 95–115.
- Savage-Rumbaugh, E. S., Rumbaugh, D. M., Smith, S. T., & Lawson, J. (1980). Reference: The linguistic essential. *Science*, 210, 922–925.
- Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus Californianus*) is capable of forming equivalence relations. *The Psychological Record*, 43, 823–839.
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14, 5–13.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.
- Sidman, M., & Cresson, O. (1973). Reading and cross-modal transfer of stimulus equivalence in severe retardation. *American Journal of Mental Deficiency*, 77, 515–523.
- Sidman, M., Cresson, O., & Willson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior*, 22, 261–273.
- Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, 43, 21–42.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5–22.
- Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities*, 6, 1–19.
- Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, 77, 556–566.
- Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using matching-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities*, 6, 41–58.
- Stromer, R., & Mackay, H. A. (1993). Delayed identity matching to complex samples: Teaching students with mental retardation spelling and the prerequisites for equivalence classes. *Research in Developmental Disabilities*, 14, 19–38.
- Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior*, 37, 329–348.
- VanBierivliet, A. (1977). Establishing words and objects as functionally equivalent through manual sign training. *American Journal of Mental Deficiency*, 82, 178–186.
- Williams, D. C., & Saunders, K. J. (1997). Methodological issues in the study of drug effects on cognitive skills in mental retardation. In N. W. Bray (Ed.), *International review of research in mental retardation* (Vol. 21, pp. 147–185). New York: Academic Press.
- Zygmunt, D. M., Lazar, R. M., Dube, W. V., & McIlvane, W. J. (1992). Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals: A methodological note. *Journal of the Experimental Analysis of Behavior*, 57, 109–117.

Received June 12, 2001  
Final acceptance May 6, 2003

## APPENDIX A

Methods for studies training novel relations with novel stimuli.

Author	Arbitrary matching to sample training					Method
	Auditory	Visual	Visual	Other	Comparisons	
Carr et al. (2000) Exp. 2		A A	B C		2	TE
de Rose, McIlvane, Dube, Galpin and Stoddard (1988) Exp. 3		A B D	B C <sup>1</sup> B		2	DP (point)
Devany et al. (1986)		A A	B C		2	TE w/prompt <sup>1</sup>
Sidman et al. (1974) PA	name	UC <sup>1</sup> UC LC	LC <sup>2</sup> UC		6	TE
Sidman et al. (1985) Exp. 1		D D A A	E F C		3	TE
Sidman et al. (1986) <sup>1</sup>	A A	B C D D	E F		3	TE <sup>2</sup>
Spradlin and Saunders (1986) Exp. 4		A B C D	E E E E		2	instruction
VanBiervliet (1977)	B		A	C <sup>1</sup> C <sup>1</sup>	5	TE <sup>2</sup>

*Note.* Training and testing conditions are listed for all subjects (S). If individual subjects had different training or testing conditions, these conditions are either noted in the *Notes* column or listed separately. Relations are listed in the order in which they were trained; uppercase letters designate arbitrary stimuli. Entries in the *Comparisons* column represent number of comparisons presented on each trial. For the *Method* column, TE = Trial and Error, DP = Delayed Prompt. *Reinforcement Conditions* are listed for test (T) and baseline (B) trials (ext = extinction, int = intermittent, diff = differential reinforcement, mix = baseline trials interspersed with test trials, mass = baseline trials reviewed during testing but massed in different blocks or sessions than probe trials, no = only test trials were presented once testing began).

## APPENDIX A

(Extended)

Arbitrary matching to sample testing					
Reinforcement conditions	Relations tested				Notes
	Symmetry	Transitivity	Equivalence	Naming	
T-ext B-mix, int	BA CA		BC, CB		tabletop; 1 S had modified training with blank comparison trials
T-ext B-mix, int	CB <sup>1</sup> BD	AC <sup>1</sup> AD			<sup>1</sup> tabletop for Subject MS only
T-ext <sup>1</sup> B-no T-diff B-mass, diff		name-LC	CA	UC, LC	tabletop <sup>1</sup> intermittent for on-task <sup>1</sup> UC = upper case letter <sup>2</sup> LC = lower case letter
T-ext B-mix, int			FE BC, CB EB		results are reported for single-node tests only
T-ext B-mix, int	CE ED FD		BC, CB EF, FE	B, C D, E, F	<sup>1</sup> relations were trained in a different order for each S <sup>2</sup> learned 2-choice before 3-choice task
T-ext B-mix, diff	EA EB EC ED		AB, BC, CD, AC, BD, AD <sup>1</sup>		<sup>1</sup> and the reverse; not clear whether all possible trials presented
T-ext <sup>3</sup> B-no	CA		BA	A, C	<sup>1</sup> nonsense-sign production (not ASL) <sup>2</sup> correction: response modeled <sup>3</sup> praise for on-task

## APPENDIX B

Methods for studies using differential outcomes procedures.

Author	Arbitrary matching to sample training					Methods
	Auditory	Visual	Visual	Other	Comparisons	
Dube et al. (1987) Exp. 1		B	B	S <sup>R</sup>	2	TE
		C	C	S <sup>R</sup>		
		S <sup>R</sup>	S <sup>R</sup>	S <sup>R</sup>		
	A	B		S <sup>R</sup>		
		B	C	S <sup>R</sup>		
Exp. 2		D	D	S <sup>R</sup>	2	TE
Dube et al. (1989)		A	A	S <sup>R</sup>	2	TE
		B	B	S <sup>R</sup>		
		C	C	S <sup>R</sup>		
		D	D	S <sup>R</sup>		
		A	B	S <sup>R</sup>		
	B	C	S <sup>R</sup>		DP/fade <sup>1</sup>	
Dube and McIlvane (1995) Exp. 1		A	A	S <sup>R</sup>	2	TE
		B	B	S <sup>R</sup>		
Exp. 2 <sup>1</sup>		A	B	S <sup>R</sup>	2	DP, fading, sample shaping
		B <sup>2</sup>	A			
		C	C <sup>3</sup>			
		A	C <sup>3</sup>			
		D	D <sup>3</sup>			

*Note.* Training and testing conditions are listed for all subjects (S). If individual subjects had different training or testing conditions, these conditions are either noted in the *Notes* column or listed separately. Relations are listed in the order in which they were trained; uppercase letters designate arbitrary stimuli. S<sup>R</sup> in the *Other* column represents outcome-specific reinforcement contingencies in which the reinforcer was food. Entries in the *Comparisons* column represent the number of comparisons presented on each trial. For the *Method* column, TE = Trial and Error, DP = Delayed Prompt. *Reinforcement Conditions* are listed for test (T) and baseline (B) trials (ext = extinction, int = intermittent, mix = baseline trials interspersed with test trials).

## APPENDIX B

(Extended)

Arbitrary matching to sample testing					
Reinforce- ment conditions	Relations tested				Notes
	Symmetry	Transitivity	Equivalence	Naming	
					S <sup>R</sup> = two different foods
T-ext B-mix, int	CB S <sup>R</sup> B, S <sup>R</sup> C	AC	S <sup>R</sup> A		
T-ext B-mix, int			DB <sup>1</sup> , DC <sup>1</sup> , AD		<sup>1</sup> these tests reversed for 1 subject
T-ext B-mix, int	BA CB	AC	CA <sup>2</sup> , BD, DB, AD, DA, CD, DC		<sup>1</sup> DP for S-PN; fade for S:JDP <sup>2</sup> not based on common reinforcer
T-ext B-mix, int			AB, BA		
T-ext B-mix, int	BA				<sup>1</sup> follow-up study with 4 Ss who did not show equivalence in Exp. 1
			AC, CA <sup>2</sup> BC, CB <sup>3</sup> AD, DA, BD, CD, DC, S <sup>R</sup> A <sup>2</sup>		<sup>2</sup> for S-ADY only <sup>3</sup> for S-ADY only; not based on stimulus-reinforcer relations

## APPENDIX C

Methods for studies involving preexisting relations.

Author	Arbitrary matching to sample training					Method
	Auditory	Visual	Visual	Other	Comparisons	
Brady and McLean (2000)	<i>name</i>	<i>object</i> object	<i>object</i> A	<i>name</i> <sup>1</sup>	3 <sup>2</sup>	TE <sup>3</sup> , blocked trials for some S
Carr et al. (2000) Exp. 1	<i>name</i>	<i>picture</i> letter D	picture picture		3	sample shaping
Gast et al. (1979)	<i>name</i> <i>name</i>	<i>quantity</i> <i>numeral</i> word name <sup>1</sup> word name <sup>3</sup>	<i>numeral</i> <i>quantity</i>		3	gradual addition of two S <sup>-2</sup>
Mackay (1985)	<i>name</i>	<i>color</i>	color	word <sup>1</sup>	6	prompt
Mackay and Ratti (1990)	<i>name</i>	<i>numeral</i> location <sup>1</sup>	<i>numeral</i>		9	exclusion
Maydak et al. (1995)	<i>name</i> <i>name</i>	<i>numeral</i> <i>quantity</i> <i>quantity</i> <i>numeral</i>	<i>quantity</i> <i>numeral</i> quantity A	sequence <sup>1</sup> sequence		
McDonagh et al. (1984) Exp. 1a		<i>price</i> N 5P <sup>1</sup> price	N <sup>1</sup> <i>price</i> price 5P		2	DP
Sidman (1971)	<i>name</i> name	<i>picture</i> word			8 <sup>1</sup> 8 <sup>1</sup>	TE <sup>2</sup>
Sidman and Cresson (1973)	<i>name</i> name	<i>picture</i> <sup>1</sup> word			8 <sup>2</sup> 8 <sup>2</sup>	TE <sup>3</sup>
Sidman et al. (1974) (Subject JC)	<i>name</i>	<i>picture</i> <sup>1</sup> picture	word		8 <sup>2</sup> 8 <sup>2</sup>	TE <sup>3</sup>
Stromer and Mackay (1993) <sup>1</sup>	name1 name2	picture1 picture1	word1 <sup>2</sup> word2 <sup>2</sup>			TE <sup>3</sup>

*Note.* Training and testing conditions are listed for all subjects (S). If individual subjects had different training or testing conditions, these conditions are either noted in the *Notes* column or listed separately. Relations are listed in the order in which they were trained; pre-existing relations are italicized; uppercase letters designate arbitrary stimuli. Entries in the *Comparisons* column represent the number of comparisons presented on each trial. For the *Method* column, TE = Trial and

## APPENDIX C

(Extended)

Arbitrary matching to sample testing					
Reinforcement conditions	Relations tested				Notes
	Symmetry	Transitivity	Equivalence	Naming	
T-ext B-no	A-object	name-A		A <sup>1</sup>	<sup>1</sup> only for S who spoke (4 of 8); <sup>2</sup> six classes; <sup>3</sup> tabletop
T-int <sup>1</sup> B-mix, int	picture-letter picture-D	name-letter name-D	letter-D D-letter		<sup>1</sup> ext on first two test trials; reinforcement on two of last four trials
T <sup>4</sup> B-mix, diff			name-quantity name-numeral same as Set 2		<sup>1</sup> Set 2: two, four, six <sup>2</sup> prompt and correction for incorrects; included sign production <sup>3</sup> Set 1: one, three, five <sup>4</sup> not reported
T-diff B-no	word-color	name-word		word	<sup>1</sup> constructed spelling
T-ext B-no	location-numeral		name-location	location	<sup>1</sup> location in a 3 × 3 matrix
T-diff B-no		numeral-sequence A-quantity A-numeral numeral-B	quantity-A numeral-A name-A quantity-B name-B	quantity numeral-A quantity numeral-B	<sup>1</sup> P = penny; N = nickel <sup>2</sup> reinforcement conditions not given
T-ext B-mix <sup>2</sup>			N-5P 5P-N	all	
T-diff B-mass, diff			picture-word, word-picture	word	<sup>1</sup> 20 classes <sup>2</sup> with correction
T-diff B-mass, diff			picture-word, word-picture	word	<sup>1</sup> pretest >80%; trained to 100% <sup>2</sup> 20 classes <sup>3</sup> several training variations
T-diff B-mass, diff	word-picture	name-word		word	<sup>1</sup> pretest 70%; trained to 95% <sup>2</sup> 20 classes <sup>3</sup> correction when needed
T-ext B-mix, ext <sup>4</sup>		name1-word2 name2-word1	word1-word2	word1, word2, picture1	<sup>1</sup> procedure simplified for clarity <sup>2</sup> visual stimuli presented as complex stimuli <sup>3</sup> included constructed spelling <sup>4</sup> 1 cent/correct after session

Note (Continued). Error, DP = Delayed Prompt. Reinforcement Conditions are listed for test (T) and baseline (B) trials (ext = extinction, int = intermittent, diff = differential reinforcement, mix = baseline trials interspersed with test trials, mass = baseline trials reviewed during testing but massed in different blocks or sessions than probe trials, no = only test trials were presented once testing began).