

## **Meta-Analysis of Pivotal Response Training for Children with Autism Spectrum Disorder**

Gulden Bozkus-Genc and Serife Yucesoy-Ozkan  
Anadolu University

*Abstract: The main purpose of this study was to review pivotal response training and examine the efficacy of pivotal response training for children with autism spectrum disorder. The other purposes of study were to (a) examine the characteristics of participants and components of the intervention in which pivotal response training was used; (b) determine the level of efficacy of pivotal response training to teach various behaviors to children with autism spectrum disorder; (c) determine whether the effectiveness of pivotal response training differed in terms of characteristics of the intervention; and (d) determine whether percentage of nonoverlapping data, percentage of nonoverlapping corrected data, and percentage of data points exceeding median were correlated. In this study, 34 single case research articles conducted with individuals with autism spectrum disorder and published in a peer-reviewed journal in between 1979–2012 were examined. Articles primarily were descriptively analyzed and then examined by use of meta-analysis. According to results, in half of the studies, treatment integrity was assessed, generalization and maintenance data were collected; in only a quarter of the studies, social validity data were collected. Pivotal response training that focused on two of the three core features of autism spectrum disorder were found effective in influencing individual outcomes. Results also indicated that percentage of nonoverlapping data, percentage of nonoverlapping corrected data, and percentage of data points exceeding median were correlated.*

One of the evidence-based interventions used to teach individuals with autism spectrum disorder (ASD) is pivotal response training (PRT). PRT (also referred to as pivotal response teaching, pivotal response treatment, pivotal response therapy, or pivotal response intervention in the literature) is a form of naturalistic behavioral intervention based on the principles of applied behavior analysis, which assume that children's impairments can

be improved with environmental manipulations such as reinforcement, consequences, and extinction (Koegel, Koegel, & Carter, 1999; Stahmer, Suhrheinrich, Reed, Bolduc, & Schreibman, 2010). PRT was developed to facilitate generalization, increase spontaneity, reduce prompt dependency, and increase motivation (Suhrheinrich, 2010). Specific components of PRT include providing clear and appropriate cues, allowing the child to choose an activity and make choices within an activity, turn-taking, interspersing maintenance tasks with acquisition tasks, reinforcing the child's attempts, responding to multiple cues, and providing contingent reinforcement directly related to the child's response (Koegel, Koegel, & McNeerney, 2001).

The first author completed this study in partial fulfillment of the requirements for a Ph.D. in special education at Anadolu University, Eskisehir, Turkey. The Anadolu University Scientific Research Projects Commission under grant number 1307E280 supported this study. The authors thank the Anadolu University Scientific Research Projects Commission for their support. The authors are grateful to the Koegel Autism Center and Dr. Ozcan Karaaslan for their help with article access, and Bayram Genc for assisting with the statistical analysis. Correspondence concerning this study should be addressed to Serife Yucesoy-Ozkan. Anadolu Universitesi, Egitim Fakultesi Ozel Egitim Bolumu, 26470 Eskisehir, TURKEY. E-mail: syucesoy@anadolu.edu.tr.

The basic premise of PRT is that changes in certain pivotal areas of behavior will trigger changes in other behavioral areas (Koegel et al., 2001). These pivotal areas are motivation, responsivity to multiple cues, self-management, and self-initiations (Koegel & Koegel, 2006, 2012; Koegel, Koegel, Shoshan, & McNeerney, 1999). According to PRT, these skills

are pivotal because they are the foundational skills enabling learners with ASD to make widespread and generalized improvements in many other areas (National Professional Development Center on Autism Spectrum Disorders, 2012).

Numerous research studies support the effectiveness of PRT. PRT improved several language functions, including speech imitation (Laski, Charlop, & Schreibman, 1988), labeling (Koegel, Camarata, Valdez-Menchaca, & Koegel, 1998), asking-questions (Koegel, Camarata, Koegel, Ben-Tall, & Smith, 1998), spontaneous speech (Laski et al., 1988), and conversational communication (Koegel, et al., 1998). PRT has been adapted for use in teaching social skills including self-initiations (Koegel, Carter, & Koegel, 2003), joint attention (Whalen & Schreibman, 2003), sociodramatic play (Thorp, Stahmer, & Schreibman, 1995), peer social interaction (Pierce & Schreibman, 1997), and academic skills (Koegel, Singh, & Koegel, 2010). PRT has also been used to decrease problem behaviors (Baker-Ericzén, Stahmer, & Burns, 2007).

Examining the PRT studies to date reveals that three reviews have been conducted, but no meta-analyses. The purpose of the first review (Masiello, 2007) was to determine the effectiveness of PRT to improve the social-emotional and communicative behavioral outcomes of children with ASD. For this purpose, 13 research studies published between 1988 and 2003 were descriptively analyzed. It was reported that PRT was effective in improving the social-emotional and communicative behavioral outcomes of children with ASD. The second review (Bozkus-Genc & Vuran, 2013) was a qualitative document analysis of the different variables in 23 studies published from 1987 to 2011 in which PRT was used to teach social skills to children with ASD. Once again, PRT was reported to be effective for teaching social skills to children with ASD. In the third review (Toper-Korkmaz & Diken, 2013), 16 studies published from 1995 to 2011 using PRT were examined descriptively. PRT was indicated as an effective method to teach expressive language skills, social skills, and play skills. Although PRT was suggested as an effective method in the aforementioned reviews, quantitative data regarding the significance of the functional relationship, which

might show the effectiveness of PRT, were not reported. Effect size calculations must be included to describe the effectiveness of an intervention in a clear and explicit way (Kazdin, 1982). Therefore, it is important to include effect size calculations in studies to examine the effectiveness of PRT in individuals with ASD.

Several computational methods have been proposed to calculate effect size in single case research (SCR) studies (Wolery, Busick, Reichow, & Barton, 2010). The most widely used methods to calculate effect sizes in SCR studies are either regression or non-regression-based (Olive & Franco, 2007; Olive & Smith, 2005; Wendt, 2009). Several studies have examined the relationship between regression and non-regression-based effect size calculations (e.g., Olive & Smith, 2005; Parker & Hagan-Burke, 2007; Wolery et al., 2010). However, researchers still debate which calculation method should be used. Percentage of nonoverlapping data (PND) is the oldest and most widely used effect size calculation method (Mastropieri & Scruggs, 1985–1986; Parker, Vannest, & Davis, 2011; Scruggs, Mastropieri, & Casto, 1987; Scruggs, Mastropieri, Cook, & Escobar, 1986); however, it does not have sufficient sensitivity (Faith, Allison, & Gormann, 1996). If there is a very high data point in the baseline level, the intervention may seem to be ineffective even though it is effective. Conversely, if there is even a very small increase in the baseline data, the intervention may appear to be effective even though it is ineffective (Faith et al., 1996). In order to eliminate these limitations of PND, percentage of nonoverlapping corrected data (PNCD) and percentage of data points exceeding the median (PEM) were developed. PNCD suggests a data correction process before calculating the PND in order to separate the possible trend from the data before the intervention (Manolov & Solanas, 2009). PEM is based on the assumption that the median best summarizes the data in the baseline level (Ma, 2006). Even though many studies have examined the PND and PEM calculations (Parker & Hagan-Burke, 2007; Wolery et al., 2010), only a limited number of them has been conducted on PNCD (Manolov & Solanas, 2009). Therefore, it is necessary to examine whether PNCD is consistent with other

methods based on overlap. The purpose of this study was to examine the efficacy of PRT. Thus, in the present study, it was sought answers to the following questions: (a) what are the characteristics of the participants and components of the intervention in which PRT was used? (b) what is the level of the effectiveness of PRT with regard to teaching various behaviors to children with ASD? (c) does the effectiveness of PRT differ in terms of the characteristics of the intervention? (d) are the PND, PNCD, and PEM effect size calculations correlated?

## Method

### *Study Identification*

A comprehensive literature review was conducted in order to identify the research studies that would be included in the meta-analysis. Five different methods were used to identify the studies: searching the related terms in the subject indexes, searching the thesis databases, footnote chasing, hand searching, and consulting with researchers. One method used to identify the studies was to search certain keywords in the subject indexes in online databases. The online databases that we searched are given in alphabetical order, as follows: Academic Search Complete, Cambridge Journals Online, Ebrary, Oxford Journals Online, Psychology and Behavioral Science, Sage Journal, Science Direct Journals, SocINDEX with Full Text, Springer LINK Contemporary, Taylor and Francis Journals, Wiley Black, and Wilson Select Plus. As articles for studies in education, psychology, and sociology most commonly appear in these databases, the abovementioned databases were preferred for the literature review. In addition, Google Scholar was used to search for studies. Keywords such as natural language paradigm (NLP), pivotal response teaching, pivotal response training, pivotal response treatment, pivotal response therapy, pivotal response intervention, and PRT and the combination of these terms with autism, initiation, self-initiation, joint attention, motivation, empathy, response to multiple cues, social skills, communication skills, and play skills were used.

The second method employed was to search

the thesis databases. Proquest Dissertations and Theses database was reviewed to obtain information not available through the other sources. The same key words as detailed in the paragraph above were used to search the online databases. Dissertation abstracts, which matched these key words, were then reviewed manually to determine whether they met the inclusion criteria. The references sections of related theses were examined. The third method for locating studies was footnote chasing (or backward chaining), which included an examination of the references in articles of related research studies. The reference lists of books and articles of interest were examined; this led to the rapid identification of the primary studies. The fourth method used to identify studies was hand searching. This procedure enabled the search for articles that had just been published and, therefore, were not yet included in the online databases. So, hand searches examined to additional studies. The fifth method used to identify articles involved consultations with researchers. After completing the search of the online databases, we requested the full text of the articles for which only summaries had been obtained from the authors or institutions (e.g., articles not found in the online databases, research reports not obtainable through other means). We emailed researchers and institutions to request papers, including research studies related to instructional practices in PRT and those conducted with children with ASD. In addition to the abovementioned methods, we reviewed the table entitled "Empirical Support for Pivotal Response Treatment" on the University of California Santa Barbara (UCSB) Koegel Autism Center web page. This table consists of the core research studies conducted from 1979 to 2010 on the pivotal area of motivation and initiations.

The literature review and the abovementioned methods yielded a pool of 69 studies. The summary of each of the studies was examined. Next, we scanned and skimmed the full text of each of the papers. Scanning and skimming revealed that four of the studies (6%) were qualitative, 11 (16%) were informative articles and reviews, and 54 (78%) were experimental studies. The experimental studies (54 research papers) were examined

in detail based on the inclusion criteria listed in the following section.

#### *Inclusion and Exclusion Criteria*

The studies to be included in this meta-analysis met the following criteria: (a) the study was published in a peer-reviewed journal from 1979 to 2012; (b) the article was written in English; (c) the participants in the study were 1–13 years of age and diagnosed with ASD; (d) the effectiveness of PRT was examined in the study; (e) one of the SCR designs was used in the study; (f) the effect of the independent variable was shown by a line graph; and (g) there was at least one data point in the baseline phase and the intervention phase.

The inclusion criteria were chosen for specific reasons. First, since the first study on PRT was published in 1979, studies conducted since 1979 were included. Second, articles written in English were included because English is the international scientific language, and it is easier to access articles written in English. Third, almost all of the studies examining the effectiveness of PRT were conducted with children with ASD in the range of early childhood to the middle school years; therefore, we included studies of children with ASD from 1 to 13 years old. Fourth, most of the studies employed a SCR design. Because the inclusion of group experimental designs with SCR designs in meta-analysis studies might generate inaccurate results. Thus, studies only conducted with SCR design were included in this study. Line graph is used in a widespread manner in the single subject designs. Because in the other graphs (e.g., bar, column) are only shown average phase values. So, these graphs do not give information especially in the event of leaps or fluctuations in the data points. For this reason, studies using line graph were included. Last, in order to calculate effect sizes and to examine the functional relationship between the baseline and intervention phases, the articles that reported at least one data point in both the baseline and intervention phases were included.

Twenty studies were excluded from this meta-analysis. Two articles that included teaching families or paraprofessionals to use PRT and examining the degree of their use of PRT were excluded. Ten articles based on

group experimental designs were excluded. One article was excluded in which the AB design did not allow for the formation of a functional relationship between the independent and the dependent variable. Two studies that used an ABC design and comparison-based alternating treatment designs were excluded. One article that did not include data points in the baseline phase was excluded. Four articles were excluded in which the effects of the independent variable were demonstrated in column chart rather than a line graph. A final total of 34 research studies published in 12 different journals were included in the meta-analysis. The list of research studies included in this meta-analysis can be obtained from the authors.

#### *Coding Procedure*

In coding procedure, first, the file of articles was created, all the articles were carefully read and summarized in a Microsoft Office Word® file. Then, a coding key and manual that explained how to code the studies were prepared by the authors following the coding systems used by Horner et al. (2005) and Odom et al. (2010). Taking into account this coding key and manual, each article was coded into a Microsoft Office Excel® file by the first author. Studies were coded in the following terms: (a) the study tag (article number, author name, year published, and the journal name); (b) participant characteristics (diagnosis, mean age, and gender); and (c) intervention characteristics (settings, instructional modification, intervener, research model, dependent variable, inter-observer reliability, treatment integrity, progress, follow-up, generalization, and social validity).

Key considerations were taken into account while the articles were coded. First, the mean age of the participants was calculated by converting the age into months (e.g., 6 years 4 months = 76 months), summing the ages of all of the participants, and dividing the sum by the number of participants; this mean value was coded into the coding key. Participant gender was coded as described in the article. However, if the gender of the participants was unclear, the entire article was scanned and statements of gender were taken into account when gender was coded. The study setting was

coded as clinic, school, home, community setting, or unspecified. The intervener was coded as expert, family, peer, or other. The research models were coded as multiple baseline designs, ABA design, or ABAB design. The dependent variables were coded as communication and interaction skills, social skills, play skills, academic skills, or multiple skills. Progress, follow-up, generalization, inter-observer reliability, treatment integrity, and social validity were coded as either available or not available.

### *Effect Size Calculations*

In order to indicate the effectiveness of the SCR designs, the baseline and the intervention phases were compared by visual analysis, and effect size was calculated. Visual analysis involves reaching a conclusion about the effects of various independent variables on dependent variables by visually examining the graphed data. Meaningful changes in the dependent variable should be apparent when displayed graphically, and emphasis is placed on the believability of the observed change behavior (Poane, Rihgdahl, Kelley, & Glover, 2011). Effect sizes were calculated using the following three methods: PND (Scruggs et al., 1987), PNCD (Manolov & Solanas, 2009), and PEM (Ma, 2006). Because a variety of SCR designs were represented in this meta-analysis, different strategies were identified in order to calculate effect size metrics. The effect sizes for the various SCR designs were calculated as follows. For the ABA design, effect size was calculated for the baseline and intervention phases for each of the participants ( $A_1$  and  $B_1$ ). For the ABAB design, effect size was calculated for the first baseline and first intervention and for the second baseline and second intervention phases for each of the participants ( $A_1$  and  $B_1$ ;  $A_2$  and  $B_2$ ). For the multiple baseline design, effect size was calculated for each of the behaviors; then, the individual scores were averaged to obtain the average score for the study.

In order to calculate the PND score, a line parallel to the horizontal axis is drawn from the highest data point in the baseline throughout the intervention phase. For appropriate and inappropriate behaviors, the number of data points above and below this

line, respectively, are divided by the total number of data points in the intervention phase, and then multiplied by 100 (Scruggs & Mastropieri, 1998). PND scores range from 0 to 100, and can be interpreted using the conventions established by Scruggs et al. (1986). PND scores  $>90$ ,  $70-79$ ,  $50-69$ , and  $<50$  represent highly effective, fairly effective, questionable, and ineffective treatments, respectively (Scruggs & Mastropieri, 1998, 2001). In order to calculate PND scores, the graphs of each of the studies were digitally saved, then enlarged and printed. In total, 295 PND scores were calculated.

PNCD is a data correction procedure to be implemented prior to calculating the PND. The main aim of the PNCD is to eliminate a possible pre-existing data trend unrelated to the introduction of the intervention (Manolov & Solanas, 2009). In order to calculate PNCD scores, a difference series is obtained by subtracting the previous data point from every data point at the baseline level ( $n_A-1$ ). The mean of these newly computed values are calculated. Having calculated the mean, the trend correction factor for each of the data points is computed by multiplying the mean of the difference series by the order of the data point. After calculating the correction factors, a data correction process is applied that includes subtracting these correction factor scores from the original data points. After this correction is performed, a PND is calculated using the new data points. No score system exists to interpret the PNCD scores. Therefore, a PND score system was used to interpret the PNCD scores in this study. Since electronic copies of 32 of the 34 articles were obtained before calculating a PNCD, hard copies of the articles were converted to soft copies by a digital scanner. The graphics of the soft copies of the articles were saved as pictures in JPEG format by using Adobe Acrobat 9 Pro software. Saved as pictures, the graphics were named according to their order of appearance in the article and given their final forms by editing them with Microsoft Office Picture Manager 2010 software. The second author performed this process. Using Plot Digitizer 2.5 for MacOS digitized the soft copy graphics. Plot Digitizer is a Java program used to digitize scanned plots of functional data (<http://plotdigitizer.sourceforge.net>). New graphics were drawn for

each of the participants and variables using the digitized data. These new graphics were visually compared with the original ones and found to be nearly identical in all cases. In total, 284 PNCD scores were calculated using the digital data. For one of the articles, the graphics could not be viewed with Plot Digitizer. Therefore, its data points could not be digitized. Consequently, for that article, a PNCD score was not calculated.

The PEM metric is computed by calculating the percentage of the treatment data points that do not overlap with the median baseline data point (Ma, 2006, 2009). In order to calculate a PEM, a horizontal line to the x-axis is drawn from the median of the data points in the baseline phase throughout the intervention phase. For appropriate and inappropriate behaviors, the number of data points above and below this line, respectively, is divided by the total number of data points in the intervention phase, resulting in the PEM score. PEM scores range from 0 to 1. PEM scores .90–1.0, .70–.89, and <.70 represent very effective, moderately effective, and questionable or ineffective treatments, respectively (Wendt, 2009; Wolery et al., 2010). PEM scores were multiplied by 100 in this study in order to obtain scores in percentages. To calculate PEM scores, the graphs of the studies were digitally saved, enlarged, and printed. In total, 295 PEM scores were calculated.

#### *Correlation Analysis*

IBM SPSS Statistics 21.0 for MacOS was used to compute Pearson's correlation coefficients in order to examine the degree and direction of the relationship among the three different effect size calculation methods. To examine the individual effects of studies rather than using the average effect size for each study, every effect size calculated was used to calculate the correlation coefficient. A total of 284 effect sizes were used to calculate the correlation coefficient. Even though 295 effect sizes were calculated in this study, since the data points in two of the figures could not be digitized and PNCD scores could not be calculated, PND and PEM scores for these figures were excluded. Thus, for each of the calculation methods, 284 effect size scores were used. We used a total of 852 effect size scores.

#### *Reliability*

In this meta-analysis, inter-rater reliability was assessed, both for the coding process and the effect size calculations, by the second author. Of all the studies, 14 (41%) were randomly selected for an inter-rater reliability assessment; the second author coded these studies. Next, the data coded by the first and the second authors were listed and every category was compared. Inter-rater reliability was obtained by a point-by-point ratio between the first author and the second author (Wolery, Bailey, & Sugai, 1988). There were 326 agreements and 34 disagreements. The inter-rater reliability was 95.1%.

To compute inter-rater reliability for the PND and PEM scores in all of the studies, the PNCD scores of 11 of the studies (30%) were chosen. Since PNCD decreases the probability of errors in mathematical calculations, only 30% of the research studies were selected to assess the inter-rater reliability of PNCD. The calculations of the first author were compared with those of the second author. For PND calculations, there were 289 agreements (97.6%) 6 disagreements. For the PEM calculations, there were 294 agreements (99.6%) and 1 disagreement. For the PNCD calculations, there were 80 agreements (100%).

#### **Results**

Results for the participant and intervention characteristics are summarized in Table 1. As shown in Table 1, most of the studies (32.3%) were published from 2009 to 2012. The majority (44.1%) was published in journals related to ASD. As listed in Table 1, there were 125 participants in total across the 34 studies. 107 (85.6%) were diagnosed with ASD, eight (6.4%) were diagnosed with pervasive developmental disorder (PDD), and 10 (8%) had comorbid disabilities with ASD. The mean age of all of the participants was 4 years, 8 months (range = 2 years, 5 months to 12 years, 8 months). In two studies (5.9%) the mean age of the participants was 1–3 years, in 21 studies (61.8%) it was 3–6 years, in eight studies (23.5%) it was 6–9 years, and in three studies (8.8%) it was >9 years. Moreover, 89 of the participants (71.2%) were male, 26 (20.8%) were female; for 10 studies (8%),

TABLE 1

**Summary of Participant and Intervention Characteristics**

<i>Participant Characteristics</i>	<i>f</i>	<i>%</i>
<b>Diagnosis</b>		
ASD	107	85.6
ASD and additional disabilities	10	8
Pervasive Developmental Disorder	8	6.4
	( <i>n</i> = 125)*	( $\Sigma$ 100)
<b>Age</b>		
0–3 years	14	5.9
3–6 years	75	61.8
6–9 years	31	23.5
9 years and over	5	8.8
	( <i>n</i> = 125)*	( $\Sigma$ 100)
<b>Gender</b>		
Male	89	71.2
Female	26	20.8
N/A	10	8
	( <i>n</i> = 125)*	( $\Sigma$ 100)
<i>Intervention Characteristics</i>		
<b>Setting</b>		
Multiple	15	44.1
Clinic	9	26.4
School	5	14.7
Home	3	8.9
Community	2	5.9
	( <i>n</i> = 34)**	( $\Sigma$ 100)
<b>Instructional Arrangement</b>		
One to one	26	76.5
Small group	3	8.8
N/A	5	14.7
	( <i>n</i> = 34)**	( $\Sigma$ 100)
<b>Intervener</b>		
Professional (researcher/clinician/therapist/teacher)	13	38.2
Parent/caregiver	8	23.5
Student (undergraduate/graduate student)	6	17.6
Peer	4	11.7
Multiple	3	9
	( <i>n</i> = 34)**	( $\Sigma$ 100)
<b>Design</b>		
Multiple baseline	29	85.3
ABAB	4	11.8
ABA	1	2.9
	( <i>n</i> = 34)**	( $\Sigma$ 100)
<b>Dependent Variable</b>		
Language and communication/interaction skills	15	44.1
Multiple skills	11	32.5
Social skills	4	11.7
Play skills	3	8.8
Academic skills	1	2.9
	( <i>n</i> = 34)**	( $\Sigma$ 100)
<b>Reliability</b>		
Inter-observer reliability	34	100
Treatment integrity	15	44.1

TABLE 1—(Continued)

<i>Intervention Characteristics</i>	<i>f</i>	<i>%</i>
<b>Maintenance, Generalization, and Social Validity</b>		
Maintenance	15	44.1
Generalization	21	61.7
Social validity	8	23.5

*f*: Frequency.

*%*: Percentage.

*n*: Number of studies.

\* Total number of subjects.

\*\* Total number of research studies.

information regarding gender was not included.

The settings were multiple in 15 of the studies (44.1%), clinic in nine (26.4%), school in five (14.7%), home in three (8.8%), and community settings in two (5.8%). PRT was presented through one-to-one instruction in 26 of the studies (76.5%), and through small group instruction in three of the studies (8.8%). In five of the studies (14.7%), no information was provided regarding the instructional arrangement. PRT was presented by professionals (clinicians, therapists, or teachers) in 13 of the studies (38.2%), by families/caregivers in eight (23.5%), by students (undergraduate or graduate) in six (17.6%), by peers in four (11.7%), and by more than one intervener in three (9%).

A multiple baseline design was used in 29 of the studies (85.3%), an ABAB design was used in four (11.8%), and an ABA design was used in one (2.9%). In 15 of the studies (44.1%), the dependent variables were communication and interaction skills. In 11 of the studies (32.5%) multiple skills were targeted. In four studies (11.7%) social skills, in three studies (8.8%) play skills, in one study (2.9%) academic skills were the dependent variables. In all of the studies (100%), PRT improved the dependent variables. In all of the studies (100%), inter-observer reliability data were gathered (*range* = 80–99). Treatment integrity was assessed in only 15 (44.1%) of the studies (*range* = 78.3–100) were assessed. Moreover, in 20 (55.8%), 15 (44.1%), and eight (23.5%) of the studies, generalization data, maintenance data, and social validity data were collected, respectively.

**TABLE 2**

**Descriptive Statistics for PND, PNCD, and PEM Scores**

<i>Calculation</i>	<i>N</i>	<i>Min.</i>	<i>Max.</i>	<i>X</i>	<i>SD</i>
PND	284	0.00	100	76.10	33.65
PNCD	284	0.00	100	78.03	34.38
PEM	284	0.00	100	89.34	22.18

*N*: Number of studies.  
*Min.*: Minimum value.  
*Max.*: Maximum value.  
*X*: Mean.  
*SD*: Standard deviation.

The effect sizes of the studies included in this meta-analysis are listed in Table 2. The table demonstrates that, for all of the studies, the mean PND score was 76.10% (*range* = 0–100%, *standard deviation* [*SD*] = 33.65), the mean PNCD score was 78.03% (*range* = 0–100%, *SD* = 34.38), and the mean PEM score was 89.34% (*range* = 0–100%, *SD* = 22.18). The level of effect for the PND, the PNCD, and the PEM scores of the studies included in this meta-analysis are listed in Table 3. As demonstrated in Table 3, with regard to PND scores, the PRT effect sizes were greater than 90% in 13 studies (38.2%), 70–89% in 11 studies, and below 70% in 10 studies (29.4%). With regard to PNCD scores, effect sizes were greater than 90% in 14 studies (41.1%), 70–89% in nine studies (26.5%), and below 70% in 10 studies (26.4%). With

regard to PEM scores, effect sizes were greater than 90% in 27 studies (79.4%), 70–89% in four (11.7%), and below 70% in three studies (8.8%). The effect size scores of every study and each calculation type can be obtained from the authors.

The PND, PNCD, and PEM scores in this study were examined in terms of different variables. The results obtained are given in Table 4. As shown in Table 4, effect size scores were greater than 70% in all of the research models. PND scores for all of the dependent variables except play skills and social skills were greater than 70%. PNCD scores for all of the dependent variables except play skills were greater than 70%.

PEM scores for all of the dependent variables were greater than 70%. For all of the studies, regardless of whether inter-observer reliability, treatment integrity, maintenance, or generalization data were gathered, all three-effect size scores were greater than 70%. In the studies in which social validity data were gathered, PND and PNCD scores were below 70% whereas PEM scores were above 90%. In studies in which researchers did not collect social validity data, all three-effect size scores were greater than 70%.

In order to examine the degree and direction of the relationship among PND, PNCD, and PEM scores, correlation coefficients were computed; they are listed in Table 5. As shown in Table 5 there were strong, positive correlations between PND and PNCD ( $r = .749, p < .001$ ), PND and PEM ( $r = .598, p < .001$ ), and

**TABLE 3**

**Level of Effect for PND, PNCD, and PEM Scores**

<i>Overlap Method</i>	<i>Level of Effect</i>					
	<i>Highly Effective (&gt;90%)</i>		<i>Fairly Effective (70-89%)</i>		<i>Questionable or Ineffective (&lt;70%)</i>	
	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>
PND	13	38.2	11	32.3	10	29.4
PNCD	14	41.1	9	26.4	10	29.4
PEM	27	79.4	4	11.7	3	8.8

*f*: Frequency.  
*%*: Percentage.



**TABLE 4**

**Means and standard deviations of the PND, PNCD, and PEM scores for the intervention characteristics**

Variables	PND			PNCD			PEM		
	N	X	SD	N	X	SD	N	X	SD
<b>Design</b>									
Multiple baseline	29	76.69	20.27	28	78.32	12.87	29	89.95	15.04
ABAB	4	91.22	15.20	4	95.48	8.53	4	98.32	2.9
ABA	1	97.2	0	1	98.2	0	1	97.2	0
<b>Dependent variable</b>									
Communication and interaction skills	13	90.27	19.25	12	90.2	13.83	13	93.16	18.1
Multiple skills	12	76.18	16.96	12	78.28	13.07	12	92.54	5.92
Social skills	5	68.48	11.96	5	70.44	8.8	5	87.58	11.9
Play skills	3	57.3	20	3	66.3	1.8	3	79.9	17.3
Academic skills	1	84.2	0	1	100	0	1	100	0
<b>Improvement</b>									
Yes	34	76.10	20.54	33	78.03	19.80	34	89.34	14.44
No	0	-	-	-	-	-	-	-	-
<b>Interobserver reliability</b>									
Yes	34	76.10	20.54	33	78.03	19.80	34	89.34	14.44
No	0	-	-	-	-	-	-	-	-
<b>Treatment integrity</b>									
Yes	15	72.41	22.6	15	73.28	9.62	15	87.06	19.14
No	19	84.21	16.37	18	87.43	13.68	19	94.38	7.01
<b>Maintenance</b>									
Yes	15	76.3	19.05	15	77.79	11.25	15	90.31	10.79
No	19	81.14	20.88	18	83.67	12.95	19	91.82	16.41
<b>Generalization</b>									
Yes	21	83.84	14.83	20	85.83	12.49	21	93.31	7.63
No	13	71.2	24.84	13	73.57	14.52	13	87.66	20.39
<b>Social validity</b>									
Yes	8	69.48	22.46	7	63.97	20.39	8	91.3	13.82
No	26	81.93	18.54	26	85.58	12.06	26	91.11	14.35

N: Number of studies.

X: Mean.

SD: Standard deviation.

PNCD and PEM ( $r = .583, p < .001$ ) (Cohen, 1988).

**Discussion**

The participant and intervention characteristics in the studies in which PRT was used were examined in this study. The effectiveness of PRT in teaching various behaviors to children with ASD was demonstrated. We assessed whether the effectiveness of PRT changed in terms of the characteristics of the intervention. We examined the relationships among

the effect sizes calculated by the PND, PNCD, and PEM scores.

When we examined the years in which the studies were published, we observed a linear increase in the number of studies conducted from 1979 to 2012. Thus, the needs of children with ASD in the basic skill areas may have become more prominent in recent years, and PRT awareness may have increased since the 2000s (Bozkus-Genc & Vuran, 2013). Most of the studies on PRT were conducted with children with ASD aged 0–9 years (91.2%). In addition, the majority of the participants were

**TABLE 5**

**Correlation Coefficients between PND, PNCD, and PEM Scores**

	<i>PND</i>	<i>PNCD</i>	<i>PEM</i>
<b>PND</b>			
Pearson's <i>r</i>		.749	.598**
Sig. (2-tailed)			
<i>N</i>	284		
<b>PNCD</b>			
Pearson's <i>r</i>	.749**		.583
Sig. (2-tailed)			
<i>N</i>	284		
<b>PEM</b>			
Pearson's <i>r</i>	.598	.583**	
Sig. (2-tailed)			
<i>N</i>	284		

PND: Percentage of nonoverlapping data.

PNCD: Percentage of nonoverlapping corrected data.

PEM: Percentage of data points exceeding median.

\*\* Correlation is significant at the .01 level (2-tailed).

male (72.2%). Children's early years are the foundation of their future. A growing body of literature suggests that early intensive intervention may greatly enhance outcomes for children with ASD (Vismara & Rogers, 2007). In addition, statistics show that autism is four to five times more common among boys than girls (Centers for Disease Control and Prevention, 2012). Therefore, it is reasonable that the majority of the participants were young (Koegel et al., 2012) male (Bozkus-Genc & Vuran, 2013; Masiello, 2007) children.

PRT was applied in various settings, such as clinics, schools, homes, community settings, or multiple settings, and by professionals, parents/caregivers, or peers. Whereas children with ASD may direct their attention to certain characteristics of stimuli in their environments, they may completely ignore others (Lovaas, Schreibman, Koegel, & Rehm, 1971). Consequently, their generalization skills might be negatively affected. It is suggested that researchers utilize multiple settings and interveners in order to increase the learning experiences of children and enable the generalization of the newly learned skills (Dunlap, Koegel, & Burke, 1981; Lovaas et al.,

1979). In studies of PRT, skills appropriate to the needs and ages of children with ASD, such as interaction skills, social skills, play skills, and academic skills, were targeted. ASD is characterized, in varying degrees, by difficulties in social interaction, verbal and nonverbal communication, and repetitive behavior. Therefore, it can be suggested that participants' core deficiency skills were targeted, which may have positively affected the social validity of the studies.

Inter-observer reliability data were collected in all of the studies. However, treatment integrity data were collected in only 44% of the studies. Accurate measurement of the dependent and independent variables is an important prerequisite to experimentally establish the existence of a functional relationship. This practice is also important for purposes of external validity and the replication of the procedures used within a study (Wheeler, Baggett, Fox, & Blevins, 2006). Therefore, future research studies are needed to demonstrate the treatment integrity of PRT. Maintenance and generalization data were collected in only half of the studies. However, like all students with disabilities, the maintenance and generalization of newly acquired behaviors are crucial for children with ASD. Moreover, maintenance and generalization are key elements to increase the instructional efficacy, allowing for the allocation of more time to teaching new behaviors. Although social validity is also critical, social validity data were collected in only one-quarter of the studies. This ratio is very low for studies in which the core skills of children with ASD were targeted (Bozkus-Genc & Vuran, 2013; Masiello, 2007). Quality indicators such as maintenance, generalization, and social validity data were not collected in some studies. The lack of these quality indicators decreased the quality of the studies (Horner et al., 2005), and may pose a limitation to the some previous studies.

This study found that PRT was effective in teaching various behaviors to children with ASD. Based on an examination of the mean effect sizes of the studies included in this meta-analysis, calculated by three methods, we conclude that PRT is fairly effective (Ma, 2006, 2009; Scruggs & Mastropieri, 1998, 2001). The current findings support previous research that individually demonstrated the

effectiveness of PRT in teaching various behaviors to children with ASD (Jones & Feeley, 2007; Koegel et al., 2012). In practice and future research studies, PRT can be used to teach pivotal behaviors to children with ASD. When the individual effect sizes of the 34 studies were examined in terms of PND scores, 13 studies were highly effective, 11 were fairly effective, and 10 were questionable or ineffective. In terms of PNCD scores, of the 33 studies, 14 were highly effective, nine were fairly effective, and 10 were questionable or ineffective. In terms of PEM scores, of the 34 studies, 27 were highly effective, four were fairly effective, and three were questionable or ineffective. When the calculation methods were compared, it was found that the PND and the PNCD scores yielded similar results, whereas these two methods were highly differentiated from the PEM scores. This result is consistent with the findings of Wolery et al. (2010). Visual analysis of the graphs of the studies in which the effect size calculations differed greatly revealed a leap or fluctuation in the data points at the baseline level; these leaps or fluctuations may have caused this difference in the calculations (Faith et al., 1996; Ma, 2006; Scruggs & Mastropieri, 1998). In the literature, researchers have mentioned that different effect size calculation methods may lead to different conclusions about the degree of treatment effectiveness for the same data set (Parker et al., 2007).

In this study, the PND, PNCD, and PEM scores were examined in terms of different variables; however, the only difference found was in terms of dependent variables. While the PND scores were questionable or ineffective in terms of play skills or social skills, PNCD scores were only questionable or ineffective in terms of play skills. The PEM scores were fairly or highly effective in terms of all of the dependent variables. This finding supports the fact that PRT was especially developed for teaching pivotal areas such as communication and interaction skills (Koegel, Koegel, Harrower, & Carter, 1999). However, it is important to consider the fact that in this meta-analysis, a few studies targeted play skills and social skills. Importantly, the data of these studies were not analyzed by inferential statistics; instead, they were analyzed by using only descriptive statistics. Inferences based on the results of a few

studies analyzed by only descriptive statistics may not be accurate.

Correlation coefficients calculated to determine the degree and the direction of the relationship among the PND, PNCD, and PEM scores indicated that there was a significant relationship among these three calculation methods; the relationship between PND and PNCD scores was stronger than the relationship between PND and PEM scores and PNCD and PEM scores (Cohen, 1988). This outcome was expected, since these three methods are based on overlap or the same theoretical foundation (Wolery et al., 2010). PNCD and PEM scores were strongly and positively correlated with PND scores, even though they were developed to overcome the limitations of calculating PND scores reported to be insufficiently sensitive. Thus, because it is easier to calculate PND scores than PEM and PNCD scores, the calculation of PND scores may be preferred for effect size calculations.

PNCD is an effective method to deal with trends. Therefore, it can be used in situations when pre-intervention measurements do not show pure random fluctuations (Malonov & Solanas, 2009). In this meta-analysis, PND and PNCD scores differed because there was a trend in the baseline for a few studies; however, this difference was not statistically significant. Therefore, when there is no trend, PND scores might be calculated instead of PNCD scores (Manolov & Solanas, 2009), since it is very difficult to digitize the data and calculate PNCD scores.

This study has some strength. First, quantitatively synthesizing SCR studies is a laudable goal because it can increase the objectivity of syntheses and allow the quantification of the potential effects (Wolery et al., 2010). Therefore, the current study may support the existing literature about PRT and teaching children with ASD. Second, PNCD is a method developed to omit pre-intervention trends from the data. Thus, PNCD can be used when the data are sequentially dependent on each other, which allows for the calculation of effect sizes using all of the data points in both baseline and intervention phases. However, PNCD scores were used in a limited number of studies. Therefore, the present study might have contributed to the related literature. Third, this study was conducted to compare

the different effect size calculations for SCR designs. Therefore, the results of the present study might provide clues and guide experts from related fields.

Several limitations need to be addressed within the framework of this meta-analysis. This study was limited to studies in which SCR designs were used because studies in which group experimental designs were used were excluded. In future meta-analyses, studies in which group experimental designs were used to test the effectiveness of PRT might be examined. In the current study, the quality of the studies included was not assessed, and the effect size calculations were not assessed in terms of the quality of the studies. However, it is crucial to examine the relationship between the quality and effect size of the studies. Therefore, new studies might be designed in order to examine the quality of the studies and the relationship between the quality and effect size of the studies. PNCD may not be a preferred method because the data points must be digitized, the calculations are very difficult, there are no criteria that can be used to interpret the scores, and it can give misleading scores when there is a trend in the data (Manolov & Solanas, 2009). Therefore, a software program might be developed to facilitate the calculation of PNCD scores. As the number of studies was limited, descriptive statistics were used in this study to examine the effect sizes in terms of the different variables. However, future analyses should examine the effect sizes of more studies by using inferential statistics in terms of different variables. The overlap methods are not an estimate of the magnitude of the effects between conditions, although they are meant to represent magnitude. The magnitude of the data (i.e., effect size) can be quite different. Overlap methods report only the proportion of overlap across the conditions. Thus, they do not provide an estimate of the magnitude of effects (Wolery et al., 2010).

## References

Baker-Ericzen, M. J., Stahmer, A. C., & Burns, A. (2007). Child demographics associated with outcomes in a community-based pivotal response training program. *Journal of Positive Behavior Interventions, 9*, 52–60.

Bozkus-Genc, G., & Vuran, S. (2013). Examining the studies in which the pivotal response instruction is used to teach social skills. *Educational Sciences: Theory and Practice, 17*, 1715–1742.

Centers for Disease Control and Prevention. (2012). Retrieved May 18, 2013, from [http://www.cdc.gov/media/releases/2012/p0329\\_autism\\_disorder.html](http://www.cdc.gov/media/releases/2012/p0329_autism_disorder.html)

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

Dunlap, G., Koegel, R. L., & Burke, J. C. (1981). Educational implications of stimulus overselectivity in autistic children. *Exceptional Education Quarterly, 20*, 37–49.

Faith, M. S., Allison, D. B., & Gorman, B. S. (1996). Meta-analysis of single-case research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 245–277). Hillsdale, NJ: Lawrence Erlbaum.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-case research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–180.

Jones, E. A., & Feeley, K. M. (2007). Parent implemented joint attention intervention for preschoolers with autism. *The Journal of Speech-Language Pathology and Applied Behaviour Analysis, 2*, 252–268.

Kazdin, A. E. (1982). Symptom substitution, generalization, and response covariation: Implications for psychotherapy outcome. *Psychological Bulletin, 91*, 349–365.

Koegel, R. L., Camarata, S., Koegel, L. K., Ben-Tall, A., & Smith, A. E. (1998). Increasing speech intelligibility in children with autism. *Journal of Autism and Developmental Disorders, 28*, 241–251.

Koegel, L. K., Camarata, S. M., Valdez-Menchaca, M., & Koegel, R. L. (1998). Setting generalization of question asking by children with autism. *American Journal on Mental Retardation, 102*, 346–357.

Koegel, L. K., Carter, C. M., & Koegel, R. L. (2003). Teaching children with autism self-initiations as a pivotal response. *Topics in Language Disorders, 23*, 134–145.

Koegel, L. K., & Koegel, R. L. (2006). *Pivotal response treatments for autism communication, social, and academic development*. Baltimore: Paul H. Brookes Publishing Co.

Koegel, R. L., & Koegel, L. K. (2012). *The PRT pocket guide, pivotal response treatment for autism spectrum disorders*. Baltimore: Paul H. Brookes Publishing Co.

Koegel, R. L., Koegel, L. K., & Carter, C. M. (1999). Pivotal teaching interactions for children with autism. *School Psychology Review, 28*, 576–594.

Koegel, L. K., Koegel, R. L., Harrower, J. K., & Carter, C. M. (1999). Pivotal response interven-

- tion I: Overview of approach. *The Journal of the Association for Persons with Severe Handicaps*, 24, 174–85.
- Koegel, R. L., Koegel, L. K., & McNERNEY, E. K. (2001). Pivotal areas in intervention for autism. *Journal of Clinical Child Psychology*, 30, 19–32.
- Koegel, L. K., Koegel, R. L., Shoshan, Y., & McNERNEY, E. (1999). Pivotal response intervention II: Preliminary long-term outcome data. *The Journal of the Association for Persons with Severe Handicaps*, 24, 186–198.
- Koegel, L. K., Singh, A. K., & Koegel, R. L. (2010). Improving motivation for academics in children with autism. *Journal of Autism and Developmental Disorders*, 40, 1057–1066.
- Laski, K. E., Charlop, M. H., & Schreibman, L. (1988). Training parents to use the natural language paradigm to increase their autistic children's speech. *Journal of Applied Behavior Analysis*, 21, 391–400.
- Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin*, 86, 1236–1254.
- Lovaas, O. I., Schreibman, L., Koegel, R., & Rehm, R. (1971). Selective responding by autistic children to multiple sensory inputs. *Journal of Abnormal Psychology*, 77, 211–222.
- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-case research. *Behavior Modification*, 30, 598–617.
- Ma, H. H. (2009). The effectiveness of intervention on the behavior of individuals with autism. A meta-analysis using percentage of data point exceeding the median of baseline phase. *Behavior Modification*, 33, 339–359.
- Manolov, R., & Solanas, A. (2009). Percentage of nonoverlapping corrected data. *Behavior Research Methods*, 41, 1262–1271.
- Masiello, T. L. (2007). Effectiveness of pivotal response training as a behavioral intervention for young children with autism spectrum disorders. *Winterberry Research Syntheses*, 14, 1–11.
- Mastropieri, M. A., & Scruggs, T. E. (1985-1986). Early intervention for socially withdrawn children. *The Journal of Special Education*, 19, 429–441.
- National Professional Development Center on Autism Spectrum Disorders (2012). *Evidence-based practices for individuals with autism spectrum disorders*. Retrieved April 4, 2013, from <http://autismpdc.fpg.unc.edu/content/evidence-based-practices>
- Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K. (2010). Evaluation of comprehensive treatment models for individuals for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 40, 425–436.
- Olive, M. L., & Franco, J. H. (2007). Effect size matters: and so does the calculation. *The Behaviour Analyst Today*, 8, 76–86.
- Olive, M. L., & Smith, B. W. (2005). Effect-size calculations and single-subject design. *Educational Psychology*, 25, 313–325.
- Parker, R. I., & Hagan-Burke, S. (2007). Median-based overlap analysis for single case data: A second study. *Behavior Modification*, 31, 919–936.
- Parker, R. I., Vannest, S., & Davis, J. L. (2011). Effect size in single-case research: A review of nine non-overlap techniques. *Behavior Modification* 35, 303–322.
- Pierce, K., & Schreibman, L. (1997). Increasing complex social behaviors in children with autism: Effects of peer-implemented pivotal response training. *Journal of Applied Behavior Analysis*, 28, 285–295.
- Poane, H. S., Rihgdahl, J. E., Kelley, M. E., & Glover, A. C. (2011). Single-case experimental designs. In W.W. Fisher, C.C. Piazza, & H.S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 132–147). New York, NY: The Guilford Press.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single case research. *Behavior Modification*, 22, 221–243.
- Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize participant research: Ideas and applications. *Exceptionality*, 9, 227–244.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education*, 8, 24–33.
- Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-case research. *Behavioral Disorders*, 11, 260–271.
- Stahmer, A. C., Suhrheinrich, J., Reed, S., Bolduc, C., & Schreibman, L. (2010). Pivotal response teaching in the classroom setting. *Preventing School Failure*, 54, 265–274.
- Suhrheinrich, J. B. (2010). A sustainable model for training teachers to use pivotal response training, Pro Quest LLC, Dissertation, University of California, San Diego, California.
- Thorp, D. M., Stahmer, A. C., & Schreibman, L. (1995). Effects of sociodramatic play training on children with autism. *Journal of Autism and Developmental Disorders*, 25, 265–282.
- Toper-Korkmaz, O., & Diken, I. H. (2013). Temel tepki öğretimi ile gerçekleştirilen etkililik araştırmalarının betimsel analizi [Descriptive analysis of effectiveness studies of pivotal response training]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi [Ankara University Faculty of Educational Sciences Journal of Special Education]*, 14, 41–55.
- Vismara, L. A., & Lyons, G. L. (2007). Using perse-

- verative interests to elicit joint attention behaviors in young children with autism: Theoretical and clinical implications for understanding motivation. *Journal of Positive Intervention*, 9, 214–228.
- Wendt, O. (2009). Calculating effect sizes for single-subject experimental designs: An overview and comparison. Paper presented at the Ninth Annual Campbell Collaboration Colloquium, Oslo, Norway.
- Whalen, C., & Schreibmen, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44, 456–468.
- Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities*, 21, 45–54.
- Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavioral analysis with exceptional students*. Boston: Allyn & Bacon Inc.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, 44, 18–28.
- 

Received: 15 December 2014

Initial Acceptance: 12 February 2015

Final Acceptance: 16 April 2015