1-1-2017

A meta-analytic review of play therapy with emphasis on outcomes measures

Scott A. Jensen
University of the Pacific, sjensen@pacific.edu

Judith N. Biesen

Elizabeth R. Graham

Recommended Citation

Jensen, S. A., Biesen, J. N., & Graham, E. R. (2017). A meta-analytic review of play therapy with emphasis on outcomes measures. *Professional Psychology: Research and Practice, 48*(5), 390–400. DOI: 10.1037/pro0000148
https://scholarlycommons.pacific.edu/cop-facarticles/609

A Meta-Analytic Review of Play Therapy with Emphasis on Outcome Measures

Scott A. Jensen, University of the Pacific

Judith N. Biesen, University of Notre Dame

Elizabeth R. Graham, Pamona College

Author Note

Scott A. Jensen received his Ph.D. in Counseling Psychology from Colorado State University. He is currently an Associate Professor and Chair of the Department of Psychology at the University of the Pacific in Stockton California. His research focuses on parenting and child behavior problems, and especially interventions to help improve child behavior.

Judith N. Biesen, received her MA in psychology from the University of the Pacific and is currently a doctoral student in clinical psychology at the University of Notre Dame. Her research focuses on the improvement of couples' interaction patterns, particularly when one or both partners experience mental health issues. She has had work published in *Journal of Family Psychology*, *Journal of Consulting and Clinical Psychology*, and *Journal of Language and Social Psychology*.

Elizabeth R. Graham received her PhD in cognitive psychology from Claremont Graduate University. She is currently the Assistant Director of Institutional Research at Pomona College.

Correspondence concerning this article should be addressed to Scott A. Jensen, Department of Psychology, University of the Pacific, 3601 Pacific Ave., Stockton, CA, 95211. E-mail: sjensen@pacific.edu

Abstract

A meta-analysis of 100 studies of outcomes of play therapy interventions was conducted to determine overall effect of intervention. Different from the Bratton et al. (2005), but more consistent with the Bratton and Lin (2015) findings, there was a significant moderate effect for play therapy interventions across all outcomes (d = .44). When four studies with aberrantly large effects were removed, this effect was d = *.36.* These moderate effects are consistent with other meta-analyses that find lower effect sizes for non-behavioral interventions. Overall study quality was poor with no studies meeting the criteria of randomized control trials. Research on play therapy interventions also does not use diagnostic criteria as part of study inclusion, thus preventing play therapy interventions from being considered as Empirically Supported Treatments. Meta-analytic findings differed by type of measure used, with measures of family functioning/relationships finding larger results than other types of measures. There was not strong consistency in measures used across the studies, with only a few measures being used across more than one or two studies. Of the more frequently used measures, the Measurement of Empathy in Adult Child Interaction (MEACI) resulted in much larger effects than other measures used, and should be further evaluated in terms of appropriate interpretation and use. Effect sizes also differed based on the reporting source, with teachers tending to rate lower impacts of treatment than other reporters.

Public Significance Statement: This study reviews 100 studies comparing Play Therapy treatment outcomes for children and adolescents and finds a small to moderate benefit for those that are treated. The average benefit for Play Therapy is about half of what has been shown previously for more behaviorally oriented treatments. Unfortunately, the overall quality of research in Play Therapy is poor compared to research standards.

A Meta-Analytic Review of Play Therapy with Emphasis on Outcome Measures

The Association for Play Therapy (2014) defined play therapy as "the systematic use of a theoretical model to establish an interpersonal process wherein trained play therapists use the therapeutic powers of play to help clients prevent or resolve psychosocial difficulties and achieve optimal growth and development" . Play therapy has an extensive history dating back to the early 1900's but was further refined to its present form in the 1950's and 60's (see Porter, 2009 for a review). Typically, a trained mental-health provider works with the child either in a group or individually. Most children are under the age of 12, however, play therapy has also been used with adolescents (Carmichael, 2006; Landreth, 2002). Play therapists employ a variety of tools to engage children in treatment, including commonly a sand tray filled with various object and miniature figures, puppets used to act out scenarios, reading and telling stories, story-making and role-playing, etc. According to the Association for Play Therapy (APT), a play therapist is a licensed mental health professional who holds either a master's level or doctoral degree, and who received extensive training and supervision in play therapy. As of January 2017, the organization recognized a total of 2707 clinicians (see Kool & Lawver, 2010, Landreth, 2002, & *Schaefer & Drewes, 2014 for more*).

Though play therapy interventions have a long history, significant research on the efficacy of such interventions has not been prevalent until the past several decades, in which there have been large increases in published research on play therapy as an intervention for children and adolescents, including several meta-analytic reviews. Notably, previous meta-analytic reviews of child interventions do not clearly identify the inclusion of play therapy interventions. The early meta-analyses of child psychotherapy outcomes, however, did establish a strong case for the effectiveness of child psychotherapy treatments with effect sizes ranging

from *d* = .71 to .79 (Casey & Berman,1985; Kazdin, Bass, Ayers, and Rogers, 1990; Weisz, Weiss, Alicke, & Koltz, 1987; Weisz, Weiss, Han, Granger, & Morton; 1995). In addition to the support for child interventions generally, however, these analyses found behavioral treatments to be much more effective than non-behavioral treatments (*d* = .91 vs. .40; Weisz et al. 1995).

Through the 1990s, work was underway to establish a process for evaluating individual treatments by disorder to determine which treatments worked for specific disorders within specific populations (Chambless & Hollon, 1998). In 1998, the Journal of Clinical Child Psychology devoted a full issue to reviews of ESTs (Lonigan, Elbert, & Johnson, 1998). As noted by Phillips (2010), absent from all of the reviews of ESTs for children were any mentions of play therapy interventions. Although play therapy interventions are widely used (Bratton & Ray, 2000), none met criteria as an EST in either the 1998 or the updated 2008 reviews.

At the time of the publication of the first reviews of EST for children, empirical evidence for play therapy was lagging behind that of other interventions, and the scientific rigor was lacking in much of the published and unpublished research on the topic (Phillips, 2010). As noted, however, the past two decades have witnessed a significant increase in the publication of research on Play Therapy outcomes, and four meta-analyses have since been published (Leblanc & Ritchie, 2001; Bratton, Ray, Rhine, & Jones, 2005; Lin & Bratton, 2015; Ray, Armstrong, Balkin & Jayne, 2015). The first meta-analysis of 42 studies (Leblanc & Ritchie, 2001) found an overall effect size of *d*=.66, as well as a positive relationship between effect size and level of parent involvement. This later finding was based primarily on the inclusion of studies examining Parent Child Interaction Therapy (PCIT; Eyberg, 1988), a behaviorally focused parent-training program, as well as filial-therapy, which trains parents to provide play therapy interventions, but often includes aspects that do not fit the traditional view of play therapy (Phillips, 2010).

Bratton et al.'s (2005) meta-analysis of play therapy included a much larger and broader range of studies (total = 93) and found an overall effect size of *d* = .80. Bratton et al. found a similarly strong effect for parent involvement (mostly based on filial therapy) as well as better outcomes for more humanistic as opposed to non-humanistic treatments. They did not include studies of PCIT in the review.

Both the Leblanc and Ritchie (2001) and especially the Bratton et al. (2005) studies have been enthusiastically received by the play therapy community as strong support for the intervention, with the second study being cited at least 176 times (100 journal articles) over a nine-year span. Only Weisz et al. (1995) was cited at a comparable frequency within the first nine years following publication (approximately 177 times), with the other three child psychotherapy meta-analyses being cited only between 26 and 37 times during the first nine years following publication.

And yet, as previously noted, even several years following the publication by Bratton et al. (2005), no play therapy interventions were included in the 2008 review of ESTs of childhood problems (Silverman & Hinshaw, 2008). In a critical review of the play therapy literature, Phillips (2010) noted several shortcomings of the play therapy meta-analyses. First, as already indicated, there seems to be a possible conflation of results based on one or two intervention types that may not (filial therapy) or clearly do not (PCIT) fit the traditional play therapy definition. A second noted deficit is a lack of attention to study quality within the meta-analyses. Although Bratton et al (2005) categorized studies by study design, they did not make reference to study quality, traditionally based on Nathan and Gorman's (2002) categorization of such. Finally, Phillips further noted a lack of rigor and contradictory findings regarding the effect of greater outcomes for longer treatments.

To address the concern of rigor and study quality, Lin and Bratton (2015) published a meta-analytic review of only studies that had a comparison group and included only studies of child-centered play therapy. This review included 52 studies but found a lower effect size of d = .47. Lin and Bratton also better classified studies based on study quality and found higher effect sizes for studies with stronger design (.58 vs. .49 vs. .24 for best, next and worst designs). Only 15 studies within their analysis met criteria for the best designs.

Ray, Armstrong, Balkin, and Jayne (2015) conducted a meta-analysis of 23 studies examining outcomes of child centered play therapy interventions in school environments. They examined outcomes by types of measure and found effect sizes ranging from d = .21 to .38. The authors did not significantly address study rigor, but did find effects for shorter treatment periods than had been reported by both Leblanc and Ritchie (2001) and Bratton et al (2005).

One area of concern across the four meta-analyses of play therapy not yet addressed is in the quality of the original outcome measures used. The importance of the reliability and validity of the measures used in meta-analysis is a less commonly considered aspect, but several researchers have suggested that measure quality is likely to have a large impact on the outcomes found. For example, in a review of meta-analytic findings, Wilson and Lipsey (2001) discovered that researcher-developed measures had one of the largest impacts on differences in effect sizes amongst all of the variables they examined. Nugent (2009) has also suggested significant differences in effect sizes based on quality of outcome measures.

Bratton et al. (2005), collapsed effect sizes by study, meaning a single mean effect size was calculated for each study regardless of how many different outcome measures were included in the study or the variability of those outcome measures. Interestingly, Bratton et al. (2005) still differentiated effect size by measure type (even though the methodology had clearly combined

dissimilar measures to calculate a single effect size per study), and found the strongest outcomes for studies that included measures of family functioning and relationships (d = 1.12). The practice of collapsing effect size results across all outcome measures in a single study prevents the possibility of examining whether overall outcomes are driven more by certain measure types (e.g., family functioning and relationships versus social adjustment), and/or by certain sources of reporting (e.g., observations versus parent report).

In contrast, Ray et al. (2015) calculated and reported effect sizes separately by outcome categories, also noting the importance of examining impact by the way it was measured. This improvement in categorizing outcome by type of measure still falls short of the intent of the current analysis to delve into the specifics of measures used, how often the same measures are used across multiple studies, and the psychometric properties and value of those measures.

The purpose of the present analysis is to further examine the literature on outcomes in play therapy, with emphasis on study quality and types of outcome measures. In particular, we estimate not only an overall effect of play therapy interventions, but also its effect on specific outcome measures (e.g., Child Behavior Checklist) and categories of outcome measures (e.g., measures of behavior). Based on the results of previous work (Bratton et al., 2005; Lin & Bratton, 2015), we predict that measures of Family Functioning/Relationships will yield the largest estimated effect size, and that higher quality studies will produce larger effect sizes than those of lesser quality. In addition, we examine whether estimated effect sizes differ based on the source of the report (e.g., parent vs. teacher report). The current analysis will evaluate outcomes from those studies included in the Bratton et al. (2005) analyses as well as more recently published work.

Method

Selection of Studies Reviewed

In addition to including the 93 studies included by Bratton et al. (2005; one excluded because it was in a different language), the current review included studies produced between 2000 and 2010, as well as additional articles found through reference lists from included articles (see figure 1). Following the procedure outlined by Bratton et al. – using the key terms *play therapy, filial therapy, family play therapy, therapeutic play,* and *play in therapy* - we searched online sources (PsycINFO, ERIC, MEDLINE, and ProQuest Dissertation Abstracts) for both published and unpublished studies of play therapy interventions. The Play Therapy Outcome Research Database of the Center for Play Therapy at the University of North Texas (http://cpt.unt.edu) served as an additional major offline resource for unpublished investigations of play therapy interventions. Similar to Bratton et al., if a published investigation was the result of a dissertation study, the dissertation results were used in the current analysis due to the more complete description of study characteristics and statistical data. If data were available in more than one published study, the study with greater data was used and the other excluded.

Of the additional articles retrieved, a total of 32 (including an additional 4 published prior to 2000) met the criteria outlined by Bratton et al. (2005). However, as we collected information and calculated effect sizes, we determined that accurate effect sizes could not be calculated for all of the studies. Of the studies originally included by Bratton et al., we were unable to estimate effect sizes in 22 articles and in the 32 additional articles not reviewed by Bratton et al., effect sizes could not be calculated for two of the studies Consequently, a total of 100 studies were included – 70 studies included in Bratton et al. (2005) and 30 additional studies. Specifically, those studies retained for final analyses (a) had used a controlled research design, (b) provided

sufficient data for computing effect sizes, and (c) had either defined the study explicitly as play therapy or, if the term "play therapy" was not mentioned, met criteria according to the Association for Play Therapy's (2014) definition of play therapy. Thirty-five of the studies were published in peer-reviewed journals and 65 were dissertations.

Data for the meta-analysis were extracted by at least two reviewers for all articles. Reliability for data extraction was very high, with inter-rater reliability being 99.6% for Treatment Outcomes, 99.7% for measure category, and 99.2% for reporting source. Any discrepancies were resolved by one of the authors – who had not done the original extraction.

Description of Study Quality Analysis

The criteria proposed by Nathan and Gorman (2002) were used to assess study quality across all manuscripts included in the review. Nathan and Gorman outline three categories of studies that collect first-hand data, based on six criteria. A Type 1 study represents the greatest quality and is characterized by a) Randomization of treatment and comparison groups, b) blinded assessment, c) clear presentation of inclusion/exclusion criteria, d) use of diagnostic criteria, e) adequate sample size for statistical power, and f) clear description of statistical methods. Randomized trials that are missing some aspect of a Type 1 study are classified as Type 2. Studies that are missing multiple or most aspects of a Type 1 study are classified as Type 3.

All studies included in the current meta-analysis were reviewed for the presence of the six criteria and then rated according to Nathan and Gorman's (2002) three types by one of the authors. A second author provided inter-rater reliability ratings on 36% of manuscripts. Overall, inter-rater reliability was 88%. Discrepancies were resolved by mutual agreement.

Description of Metanalytic Method

The 100 studies that met our inclusion criteria contained a total of 765 measures. Morris's (2008) equations 6 – 8 were used to calculate an effect size, *d*, for each measure. The equations, shown below, are appropriate for pretest-posttest, control group designs and follow the same method used by Lin and Bratton (2015), including the bias correction proposed by Hedges and Olkin (1985). However, this method of estimating *d* differs from Bratton et al. (2005) who calculated effect sizes using posttest scores only.

$$d = c_P \left[\frac{\left(M_{post,T} - M_{pre,T}\right) - \left(M_{post,C} - M_{pre,C}\right)}{SD_{pre}} \right]$$

$$SD_{pre} = \sqrt{\frac{(n_T - 1)SD_{pre,T}^2 + (n_C - 1)SD_{pre,C}^2}{n_T + n_C - 2}}$$

$$c_P = 1 - \frac{3}{4(n_T + n_C - 2) - 1}$$

In the event that means and standard deviations were not available, alternate methods of estimating *d* were used (e.g., change scores, mean squared error, *t* values) when possible. Of the 765 measures, 5 effect sizes from 2 separate studies were not calculated because the authors administered those measures posttest-only.

The first meta-analysis estimated the overall treatment effect. Following Bratton et al. (2005), we calculated one effect size for each article by averaging the effect sizes of the individual measures. Thirty-six of the articles included both subscales and total scores for at least one composite measure [e.g., Parental Stress Index (total), Parental Stress Index (child domain), and Parental Stress Index (parent domain)]. When both subscales and total composite scores were reported, we included only the effect sizes for the total scores in our average. Three studies reported different sample sizes, depending on the outcome measure. In these instances,

the largest sample sizes were used to calculate the inverse variance weight for the articles' average effect sizes.

A second series of meta-analyses examined treatment effects for the five most frequently used outcome measures. These were the Child Behavior Checklist (CBCL; $n_{articles}$ = 24), Filial Problem Checklist (FPC; $n_{articles}$ = 10), Measurement of Empathy in Adult Child Interaction (MEACI; $n_{articles}$ = 9), Porter Parental Acceptance Scale (PPAS; $n_{articles}$ = 15), and the Parental Stress Index (PSI; $n_{articles}$ = 16). Most of these measures are composites of subscales, and when results for subscales and total scores were both reported, we included only the total score in our estimate of effect size. In one article, only subscales for the CBCL were reported, and these effect sizes were averaged to create a total score.

For the third series of meta-analyses, each outcome measure was coded based on measurement category. This coding scheme produced eight categories of measures: behavior, social adjustment, personality, self-concept, internalizing, family functioning/relationships, developmental/adaptive, and other. Seventy-six of the articles included measures from multiple categories.

A fourth and final series of meta-analyses was conducted based on reporting source. This coding scheme produced six categories of measures: parent-report, child-report, teacher-report, observation, test of child, and other. Fifty-six of the articles included measures from multiple categories.

For all of the analyses listed above, Field and Gillett's (2010b) SPSS syntax file, "Meta_Basic_d.sps" was used to combine effect sizes using a random effects model. Possible publication bias was evaluated using moderator analysis (Field & Gillett, 2010c) and the fail safe N (Rosenthal, 1979); possible outliers were identified using funnel plots.

Results

Study Quality

Overall, the quality of studies included in the analysis was poor, with none of the studies meeting Nathan and Gorman's (2002) criteria as a Type 1 study, 21% meeting criteria as a Type 2 study, and the remaining 79% meeting criteria as Type 3 studies. In regards to specific categories of study quality, the percentages were as follows: 1) randomization = 58%, 2) blinded assessment = 30%, 3) clear inclusion/exclusion criteria = 66%, 4) used diagnostic criteria = 2%, 5) adequate sample size for statistical power = 10%, and 6) clearly described statistics = 93%.

Meta-Analysis

The average effect sizes from the 100 included articles produced an estimated treatment effect size of $d = .44$, 95% CI: .32 - .57, $z = 6.92$, $p < .001$, suggesting a moderate effect of treatment similar to that reported by Lin and Bratton (2015) in their updated and more restricted meta-analysis. This estimated effect size corresponds to a "Number Needed to Treat" (NNT) of 4.09 which represents the number of participants needed in the treatment and control groups in order to produce an expected difference of one more successful outcome in the treatment group, compared to the control group (Kraemer & Kupfer, 2006). In the context of the current analysis, a successful outcome would be a greater pre-post change for a treatment participant, compared to a control participant. The resulting funnel plot was reasonably symmetric with the exception of four studies reporting large standard errors and effect sizes greater than 2. However, excluding these four potential outliers, and rerunning the meta-analysis with $k = 96$ studies still produced an effect size, $d = .36$, that was significantly greater than zero. There was no evidence of publication bias: The fail-safe N was 6796, and a moderator analysis found no significant

difference in effect size between published articles (d = .49) and dissertations (d = .42), χ^2 = .19, p = .66.

A moderator analysis of study quality also resulted in a non-significant difference, χ^2 = .21, p = .15. However the trend was for Type 2 studies to have a larger estimated effect size, d = .65, compared to Type 3 studies, d = .39.

The estimated effects of treatment for the five most frequently administered measures are shown in Table 1. The estimated effect size for each of these outcome measures is significantly greater than zero. One of the measures (MEACI) yielded an extremely large effect size. Bratton et al. (2005) reported excluding a commonly used measure (they did not identify which one) because it yielded particularly large effect sizes. Based on these results we suspect it was the MEACI.

Examination of the funnel plots revealed reasonably symmetric, funnel-shaped distributions for the FPC, PPAS, and PSI. The plot for the CBCL was symmetric except for two outliers. Excluding those two outliers yielded an estimated effect size significantly greater than zero, d = .28, 95% CI: .06-.49, z = 2.55, p = .01. The funnel plot for the MEACI was neither symmetric nor funnel-shaped, but suggested a strong positive linear relationship between standard error and effect size, r = .72, p < .001.

It is important to note that the effect sizes presented in Table 1 are not independent of one another. For example, seven of the nine studies that administered the MEACI also administered both the PPAS and the PSI. Similarly, nine of the sixteen studies that administered the PSI also administered the CBCL. Thus, the participants from a single given study may affect several of these outcome measures.

The estimated effects of play therapy for the different types of outcome measures are shown in Table 2. All but two measure types, Personality and Social Adjustment, yielded moderate effect sizes significantly greater than zero. Funnel plots produced the expected symmetric, funnel-shaped distribution for five of the eight measure types: Self-Concept, Developmental/Adaptive, Social Adjustment, Personality, and Miscellaneous. The distributions for Behavior and Internalizing were both symmetric with single outliers. In both cases, exclusion of the outlier had little impact on the estimated effect sizes. The distribution for family functioning/relationships was neither symmetric nor funnel-shaped, but showed a positive linear relationship between standard error and effect size, $r = .29$, $p = .02$. Because this pattern was similar to that found for the MEACI, a measure included in the family functioning/relationships category, we re-ran the analysis after excluding all of the MEACI scores. The resulting effect size was smaller, $d = .51$, but still significantly different from zero. Excluding the MEACI scores also resulted in a non-significant linear relationship between standard error and effect size, $r = .13$, $p = .29$.

Effect sizes for the different sources of reporting are shown in Table 3. All reporting sources yielded significant effect sizes. Only two of the reporting sources produced symmetric funnel plots: Teacher and test of child ability/characteristic. The funnel plots for parent and child reporters were roughly symmetric but for single outliers. Exclusion of the outlying data points had relatively minor effects on the estimated effect sizes. The distribution for Observation showed a strong positive linear relationship between standard error and effect size, $r = .50$, $p = .001$. Because the MEACI measure is produced using this reporting type, we re-ran the analysis after excluding all of the MEACI scores. The resulting effect size was smaller, see Table 3, but still significantly different from zero. Excluding the MEACI scores also produced a symmetric,

funnel shaped distribution. Finally, the distribution for "other" reporting type appeared asymmetrical, possibly with two outliers. However the small number of studies included in this category, $k = 7$, prevented further evaluation.

Discussion

The purpose of the current meta-analytic review was to update the literature on the effectiveness of play therapy with particular emphasis on study quality and type of outcome measures. Notably, the current review found that a fairly large number of studies (24%) included in the Bratton et al. (2005) meta-analysis did not have sufficient data to calculate an accurate pre-post effect size. This difference in included studies likely contributed to the large discrepancy in findings. Using pre-post methodology for calculating effect sizes, Lin and Bratton (2015) reported an estimated effect size similar to the current review.

The overall findings of the present meta-analysis suggest that play therapy interventions demonstrate a small to moderate effect ($d = .44$, NNT = 4.09). However, this estimate may be inflated by a small number of studies (n = 4) with particularly large effect sizes. When these four studies were removed the effect size was $d = .36$, NNT = 4.98. Three of these four studies included the MEACI, a measure developed specifically for play therapy. Our analysis of MEACI outcomes suggests that it tends to produce especially large effect sizes. The fourth study used an outcome measure that was very similar to the treatment, possibly explaining the large effect. The 18% decrease in overall effect with the removal of the four studies is notable and we think suggests the smaller effect size of $d = .36$ to be a more accurate representation of the overall literature.

In contrast to the conclusion drawn by Bratton et al. (2005) that play therapy interventions have comparable outcomes to those of other child-focused treatments, the current

findings suggest smaller overall effects for play therapy interventions. These findings are similar to distinctions noted in the previous reviews demonstrating advantages for behavioral treatments compared to non-behavioral treatments; for example, Weisz et al. (1995) found an advantage of d = .91 (NNT = 2.08) versus d = .40 (NNT = 4.49) for behavioral treatments vs. non-behavioral treatments. Indeed, the current findings of d = .36 or d = .44 are similar to the d = .40 effect size of non-behavioral treatments.

Regarding study quality, there were no studies available for inclusion that met Nathan and Gorman's (2002) criteria for randomized control trials, the gold-standard for establishing ESTs. Further, only 20% of the studies met the second tier of quality, with 80% of the studies representing the poorest quality. These percentages differ somewhat from those of Lin and Bratton (2015) because they used a less stringent standard of quality based on only two items: treatment integrity (use of manual, description of procedure, and training for therapist) and randomization. Similar to Lin and Bratton, there was a small trend toward high effect sizes for studies with better quality. This suggests a significant need to improve study quality within the field of play therapy. There is greater need for randomization, clear inclusion/exclusion criteria, blinded, assessment, and adequate sample size, and, if to be considered as specific treatments, much greater need for categorizing by diagnostic criteria.

Notably, only 2% of studies incorporated diagnostic criteria. Without targeting interventions at specific diagnostic criteria, treatments cannot be considered as ESTs. This likely explains, in part, the absence of play therapy interventions within reviews of ESTs. Though Lin and Bratton (2015) noted that "global behavior problems", "internalizing behavior problems", and "externalizing behavior problems" were the most common presenting issues, the current findings suggest that diagnostic criteria are rarely used as inclusion criteria. Play therapy has

historically been considered transdiagnostic, or in other words designed to treat more generic problems without addressing specific diagnoses. This is likely the reason that most studies do not address specific diagnosis, but the broader impact is an isolation of play therapy as an intervention and the prevention of discussion of play therapy within the EST literature. The generic treatment approach could be maintained while still assessing its impact on specific disorders in order to establish its viability as an EST.

In addition to clarifying the relative effectiveness of play therapy interventions, the present analysis further adds to the literature by demonstrating that type of measure, source of report, and quality of measure have a small to moderate impact on meta-analytic findings. By analyzing by measure without collapsing within study, we demonstrated that measures within the same study vary in strength and that specific measures can often drive higher effect sizes – just as was suggested by Wilson and Lipsey (2001). The MEACI, for example had an effect size more than three times larger than the next closest measure, and more than seven times larger than the overall effect. The analysis also suggests that there is inconsistency in the types of measures used to assess outcomes in play therapy interventions. The most commonly used measure (CBCL) was only used in 24% of all studies. The next most commonly used measures were included in 9-16% of studies, and most were used in only one or two studies. Previous meta-analyses do not provide sufficient information on the variety of measures used to compare if there are differences between the diversity of measures used in research on play therapy versus other intervention techniques, but the issue is worth further exploration. Though they did not report on the specifics of measure variability within categories of outcome variables, Ray et. al (2015) did find similar variability across different categories of measures.

We caution against too strong an interpretation of the current findings due to the large variety of measures used. Our analyses estimated the effect sizes associated with play therapy by grouping disparate outcome measures. Although this allowed us to collapse across studies using relatively few common outcome measures, the impact that this may have on the meaning and reliability of the combined effect sizes is unclear. An example of this issue is the combining of several of the more frequently used measures, the PSI, PPAS, and MEACI, along with other measures used only in a single study, into the category of "Family Functioning/Relationships," to produce an estimated effect size of $d = .68$. However the PSI, PPAS, and MEACI themselves yielded substantially different estimated effect sizes, ranging from $d = .61$ to 3.10. Collapsing across so many differing measures may represent a combining of metaphorical apples and oranges. Further exploration into the impacts of analyses combining across so many disparate measures is needed.

Analysis by measure type suggests that play therapy has its greatest impact on family relationship as reported by parents, and least impact on social adjustment and internalizing symptoms. Although Bratton et al. (2005) also found that play therapy had the greatest impact on family relationships, our estimated effect size is two-thirds the size of theirs (ds = .68 and 1.12). This discrepancy is likely due to two key methodological differences. The first is the differing inclusion criteria, described previously. The second key methodological difference is that Bratton et al. (2005) coded each article for the presence or absence of each type of outcome measure (e.g., behavior), but then used the average of all measures in the article, even those of different types, when calculating the effect size. By contrast, in the current review we calculated separate effect sizes for each measure type, and included only the effect size for the specific measure type when comparing across type of measure.

The variability of effect size based on measure type suggests that play therapy might be best at addressing issues of family relationships, given that this was the largest effect. One note of caution is that at least some of this effect could be explained by possible placebo/reporter effects given that parents were the most common reporting source for measures of family relationships (60%). Play therapy interventions had the least impact on measures of social adjustment and internalizing problems, so it could be considered less effective for these problems. Effects across other types of measures were similar, falling in the small to moderate range.

Regarding source of reporter, findings across most reporters are similar, falling in the small to moderate range. This does not include the impact of the MEACI on observational measures, which has been discussed previously. Teachers, however, do not report meaningful differences as a result of play therapy interventions. This variability in reporting by parents vs. teachers, is not uncommon (see Achenbach, McConaughy, & Howell, 1987; Epkins & Myers, 1994; Kendell, Kortlander, Chansky, & Brady, 1992), but should be explored further within the context of play therapy interventions.

The present meta-analytic findings confirm that play therapy interventions demonstrate a significant, small effect on a variety of outcomes. These effects are comparable to previous studies of non-behavioral interventions, but notably lower than comparative studies of behavioral interventions. Greater clarity is needed on purposes and intent of play therapy interventions. To be considered as an EST, diagnostic criteria would need to be used in future studies. Play therapy researchers should also continue to emphasize greater use of larger sample sizes, blinded assessment procedures and randomization. Such emphasis on study quality will continue to clarify the comparative usefulness of play-therapy interventions compared to other alternatives

within the child treatment literature. Additional factors that remain to be adequately explored within the play therapy literature include the impact of the interventions across various age groups, as well as further clarification of the impact of length of treatment, which was noted by Phillips (2010) as an area of concern. Ray et al. (2015) found positive effects for shorter treatments, but this was within school treatments only and contradicted previous findings. Finally, given the discrepant findings in terms of effect sizes for play therapy vs. behaviorally oriented treatment, more head to head comparisons of play therapy interventions with behaviorally focused interventions are warranted to determine best possible treatments.

References

Association for Play Therapy. (2014). Why play therapy? Retrieved from http://www.a4pt.org/?page=WhyPlayTherapy

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/Adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213-232.

Bratton, S. C., Ray, D., Rhine, T., & Jones, L. (2005). The efficacy of play therapy with children: A meta-analytic review of treatment outcomes. *Professional Psychology: Research and Practice, 36,* 376 –390. doi:10.1037/0735-7028.36.4.376

Bratton, S., & Ray, D. (2000). What the research shows about play therapy. *International Journal of Play Therapy, 9,* 47–88. doi:10.1037/h0089440

Carmichael, K.D. (2006). Legal and ethical issues in play therapy. *International Journal of Play Therapy, 15,* 83-99.

Casey, R.J. & Berman, J.S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin, 98(2),* 388-400.

Chambless, D.L. & Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66(1),* 7-18.

Epkins, C.C. & Myers, A.W. (1994). Assessment of childhood depression, anxiety, and aggression: Convergent and discriminant validity of self-, parent-, teacher- and peer-report measures. *Journal of Personality Assessment, 62,* 364-381.

Eyberg, S. M. (1988). Parent-child interaction therapy: Integration of traditional and behavioral concerns. *Child & Family Behavior Therapy, 10,* 33–46.

Field, A. P. & Gillett, R. (2010a). How to do a meta-analysis. British Journal of Mathematical

and Statistical Psychology, 63, 665-694.

Field, A. P. & Gillett, R. (2010b). Meta_Basic_d.sps [SPSS command syntax]. Retrieved from www.statisticshell.com/meta_analysis/how_to_do_a_meta_analysis.html

Field, A. P. & Gillett, R. (2010c). Meta_Mod_d.sps [SPSS command syntax]. Retrieved from www.statisticshell.com/meta_analysis/how_to_do_a_meta_analysis.html

Glass, G., McGraw, B., & Smith, M. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

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

Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58,* 729–740. doi:10.1037/0022-006X.58.6.729

Kendell, P.C., Kortlander, E., Chansky, T.E., & Brady, E.U. (1992). Comorbidity of anxiety and depression in youth: Treatment implications. *Journal of Consulting and Clinical Psychology, 60,* 869-880.

Kool, R., & Lawver, T. (2010). Play therapy: Considerations and applications for the practitioner. *Psychiatry, 7(1), 19-24.*

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry, 59*, 990-996. doi: 10.1016/j.biopsych.2005.09.014

Landreth, G.L. (2002). Therapeutic limit setting in the play therapy relationship. *Professional Psychology: Research and Practice, 33,* 529-535.

LeBlanc, M., & Ritchie, M. (2001). A meta-analysis of play therapy outcomes. *Counseling*

Psychology Quarterly, 14, 149–163. doi:10.1080/09515070110059142

Lin, Y.W., & Bratton, S.C. (2015). A meta-analytic review of child centered play-therapy approaches. *Journal of Counseling and Development, 93,* 45-58. DOI: 10.1002/j.1556-6676.2015.00180.x

Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children. [Special Issue]. *Journal of Clinical Child Psychology, 27*(2).

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6: e1000097. doi:10.1371/journal.pmed.1000097.

Morris, S. B. (2008). Estimating Effect Sizes from Pretest-Posttest-Control Group Designs. *Organizational Research Methods, 11,* 364-386. doi: 10.1177/1094428106291059

Nathan, P.E. & Gorman, J.M. (2002). *A Guide to Treatments that Work, 2nd Ed.* New York: Oxford University Press

Nugent, W.R. (2009). Construct validity invariance and discrepancies in meta-analytic effect sizes based on different measures. *Educational and Psychological Measurement, 69(1), 62-78. DOI* 10.1177/0013164408318762

Phillips, R.D. (2010). How firm is our foundation? Current play therapy research. *International Journal of Play Therapy, 19,* 13-25.

Porter, M. (2009). Play therapy: A review. *Early Child Development and Care, 179,* 1025-1049.

Ray, D.C., Armstrong, S.A., Balkin, R.S, & Jayne, K.M. (2015). Child-centered play therapy in the schools: Review and meta-analysis, *Psychology in the Schools, 52,* 107-123.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 86, 638–641. http://dx.doi.org/10.1037/0033-2909.86.3.638

Schaefer, C. E., & Drewes, A. A. (Eds.). (2014). *The therapeutic powers of play: 20 core agents of change (2nd ed.)*. Hoboken, NJ: John Wiley & Sons.

Silverman, W.K & Hinshaw, S.P. (2008). The second special issue on evidence-based psychosocial treatments for children and adolescents: A ten-year update. *Journal of Clinical Child & Adolescent Psychology, 37(1), 1-7.* DOI: 10.1080/15374410701817725

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760.

Smith, M. L., Glass, G. V, & Miller, T. I. (1980). *The benefits of psycho- therapy.* Baltimore, MD: Johns Hopkins University Press.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542–549. doi:10.1037/0022-006X.55.4.542

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117,* 450–468. doi:10.1037/0033-2909.117.3.450

Wilson, D.B. & Lipsey, M.W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6(4), 413-429.*

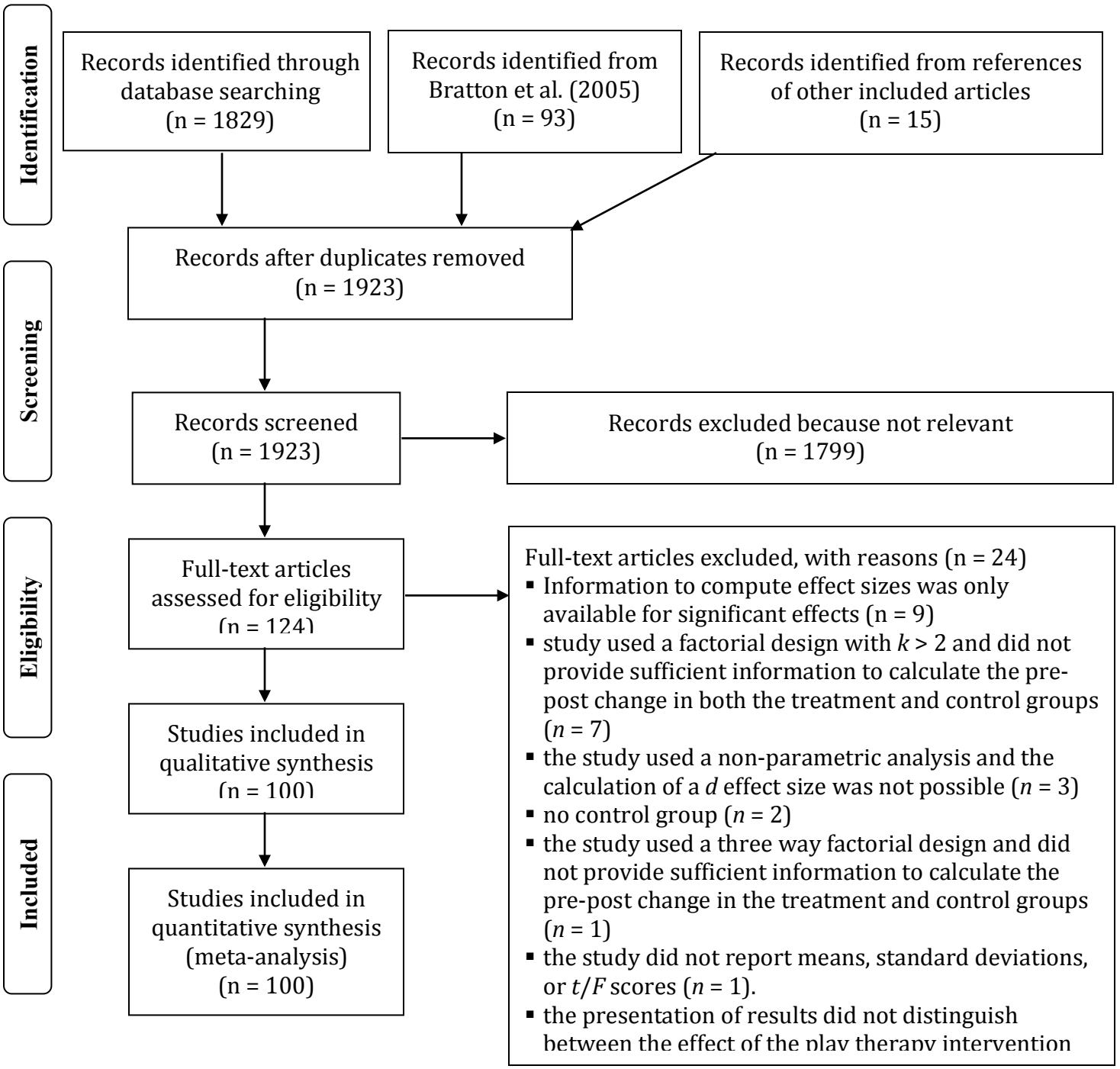

Figure 1. Adapted PRISAM 2009 flow diagram of the included studies.

Table 1. Meta-Analytic Results for the Effect of Play Therapy on the Child Behavior Checklist, Filial Problems Checklist, Measurement of Empathy in Adult Child Interaction, Porter Parental Acceptance Scale, and the Parental Stress Index.

Measure	*k*	*d*	95% CI	*SE*	*z*	*p*	Fail Safe *N*
CBCL	24	0.43	.14/.71	0.15	2.93	.003	304
PSI	16	0.61	.32/.90	0.15	4.11	< .001	251
PPAS	15	0.95	.52/1.38	0.22	4.33	< .001	545
FPC	10	0.44	.13/.76	0.16	2.74	.006	50
MEACI	9	3.10	2.34/3.85	0.38	8.07	< .001	2168

Table 2. Meta-Analytic Results for the Effect of Play Therapy on Different Types of Outcome Measures.

Type of Measure	*k*	*d*	95% CI	*SE*	*z*	*p*	Fail Safe *N*
Behavior	67	0.34	.21/.48	0.07	4.96	<.001	1850
Internalizing	42	0.23	.02/.43	0.10	2.19	.03	203
Family Functioning/ Relationships	34	0.68	.40/.96	0.14	4.83	<.001	1556
Self-Concept	30	0.35	.17/.53	0.09	3.80	<.001	352
Developmental/ Adaptive	28	0.33	.20/.46	0.07	4.92	<.001	195
Social Adjustment	14	0.12	-.09/.34	0.11	1.13	.26	-4
Personality	11	0.34	-.03/.71	0.19	1.82	.07	33
Miscellaneous	7	0.51	.28/.74	0.12	4.40	<.001	37

Table 3. Meta-Analytic Results for the Effect of Play Therapy on Different Sources of Reporting.

Source of Reporting	*k*	*d*	95% CI	*SE*	*z*	*p*	Fail Safe *N*
Parent-report	57	0.41	.23/.58	0.09	4.57	<.001	1685
Child-report	29	0.34	.05/.63	0.15	2.28	.02	194
Teacher-report	39	0.14	.03/.25	0.06	2.45	.014	37
Observation	23	1.31	.84/1.77	0.24	5.51	<.001	2979
Observation not including MEACI	11	0.53	.18/.88	0.18	2.96	.003	127
Test of child ability/characteristic	29	0.27	.09/.46	0.09	2.90	.004	154
Other	7	0.43	.15/.70	0.14	3.05	.002	15