

Animal-Assisted Therapy for Youth: A Systematic Methodological Critique

Dana. K. May¹, Nicholas P. Seivert¹, Annmarie Cano¹,
Rita J. Casey¹, & Amy Johnson²

¹Department of Psychology, Wayne State University, Detroit, MI 48201
²Teacher's Pet: Dogs and Kids Learning Together and Oakland University
Animal Assisted Therapy Certificate Program

Animal Assisted Therapy (AAT) for youth has the potential to benefit both physical and mental health outcomes. Yet little is known about the extent to which study designs in this area are aligned with established standards of intervention research. This critical review assesses current research methodologies focusing on AATs for youth with physical and mental health concerns. The main aims of this review are to advance the knowledge base of empirically supported treatments and identify next steps that researchers can take to secure the place of AATs as sound and valid interventions for youth.

Keywords: Animal assisted therapy, AAT, review, youth, children

Correspondence should be addressed to: Dana May, M.A., Department of Psychology, Wayne State University, 5057 Woodward Ave., 7th Floor, Detroit, MI 48303, phone: (313) 577-2800, fax: (313) 577-7636, email: dana.may@wayne.edu.

Introduction

Although anecdotal support for Animal-Assisted Therapy (AAT) for youth is widespread and increasing among some practitioners, it is critical that AAT has clear empirical support to be accepted in the larger fields of medicine and psychology (Kazdin & Weisz, 1998). Empirical evidence provides clinicians with needed understanding of expected outcomes and comparison to other treatment options. Aligning research of AAT with empirically supported treatment is not a new goal in the field (Kazdin, 2010); however, to our knowledge there are no reviews of the literature of AAT for youth that systematically examine the quality of its research methodology and empirical support for these interventions. Empirically supported treatments can be tested using a prescribed set of principles; these have been applied generally to AAT (Kazdin, 2010). However, current reports of these

interventions typically have small, unrepresentative samples and no control group (Krueger & Serpell, 2010; Herzog, 2011). A recent review of randomized-controlled AAT trials revealed only eight studies that met the criteria for inclusion, demonstrating the lack of rigorous methods used in this area. These authors were not able to make strong conclusions regarding efficacy of AATs due to this small sample size (Maujean, Pepping, & Kendall, 2015). The prescribed principles for empirically evaluating treatments include rigorous assessment of design, procedure, assessment, and outcome analysis (Kazdin, 2010; Kendall, Holmbeck, & Verduin, 2004).

Evaluation of study *design* involves assessment of the theoretical basis of the work. Specifically, well-designed studies present rationale for the treatments given to participants, their diagnoses to be addressed, and methodological considerations such as the inclusion of appropriate control groups to

rule out alternative hypotheses, random assignment to treatment conditions, and the timing of post-treatment follow-up time points to measure longevity of treatment effects, if they exist. An evaluation of the study *procedure* would focus on the replicability of the study design. That is, the procedure considers the participant sample including descriptive information about the sample, a clear description of the treatment, and attempts at documenting and ensuring fidelity of treatment delivery across time. Evaluation of the *assessment* procedure examines the validity, reliability, and appropriateness of the measures for the population and constructs of interest. Finally, evaluation of study *outcome analysis* is concerned with appropriateness of data analysis methods for outcome data and factors within the treatment that contribute to the outcomes (e.g. testing of mediators/moderators). This prescribed approach to evaluating therapies assures that studies are evaluated based on their scientific rigor. This review explores these characteristics and processes in the extant literature on AAT for youth. These aspects of study methodology will be evaluated with the aim of highlighting methodological strengths and weaknesses and offering directions for researchers who would like to pursue high quality AAT research in youth.

Method

Literature Review

Studies were included if they investigated the extent to which AATs had measurable benefits for youth, ages 0 to 18 years. The definition of AAT used to identify studies was a therapeutic interaction that utilizes the human-animal bond in goal-directed interventions as an integral part of the treatment process (Pet Partners, 2012). Studies published prior to December 31,

2014, were located using two methods. First, electronic searches of several databases were conducted (Medline, PsychINFO, Google Scholar) using a combination of search terms: animal-assisted therapy, AAT, animal interventions, youth, children, adolescents, dog, horse, dolphin, equine-assisted therapy, and hippotherapy. Second, references cited in relevant publications were explored to identify additional relevant articles. Studies that included unstructured time with an animal or had results that were only anecdotal or qualitative were excluded. Physical, physiological, psychological, and social outcome measures were included. Initial searches yielded approximately 4,500 studies, which were reviewed further to determine if the study met review criteria. When it was unclear if a study clearly meets the criteria for review, at least two of the current review's authors examined the methodology to determine if the inclusion criteria were met. Forty-five studies met these criteria and were included in the quality coding procedures described below. See Table 1 for a description of the 45 studies included in this review.

Table 1: Summary of Included Studies Examining Animal Assisted Therapy for Youth

Citation	Type of AAT	# of Sessions	Sample	Sample Size	Control Group	Dependent Variable
Anderson & Olson, 2006	AAT with dogs	37	Students in self-contained classroom with severe emotional disturbances	6	None	Emotional Stability
Bachi, Terkel, & Teichman, 2012	Equine Facilitated Psychotherapy	14-29	Adjudicated adolescents	29	Alternative Treatment	Self-image Self-control Trust Life satisfaction
Balluerka, Muela, Amiano, & Caldentey, 2014	AAT with dogs and horses	34	Adolescents in residential childcare	58	Treatment as usual	Attachment
Bass, Duchowny, & Llabre, 2009	Therapeutic horseback riding	12	Children with ASD	34	None	Social responsiveness Sensory profile Attention
Bertoti, 1988	Therapeutic horseback riding	11	Children with cerebral palsy	11	Repeated measures	Posture
Boyer & Mundschen, 2014	AAT with cat	12	Children with language impairments	3	Crossover design	Verbal communication Social skills
Braun, Stangler, Narveson, & Pettingell, 2009	AAT with dog	1	Pediatric inpatients with pain	57	No dog	Blood pressure Respiratory rate Pulse Pain level
Breitenbach, Stumpf, Fersen, & Ebert, 2009	Dolphin-assisted therapy	Unknown	Children with disabilities who have limited communication abilities	118	1)outpatient therapy and interactions with dolphins (no AAT), 2)interactions with farm animals only, 3)no animals and not treatment	Social-emotional competence Sociability Self-confidence Responsiveness Unstable acting-out behavior Relationship harmony Tension/relaxation Content harmony Stereotypic behavior
Cuyper, DeRider, & Stranbeim, 2011	Therapeutic horseback riding	16	Children with ADHD	5	Repeated measures	Health-Related Quality of Life Motor Skills
Dietz, Davis & Pennings, 2012	AAT with dog	12	Youth from Child Advocacy Center with childhood sexual abuse	153	Alternative Treatment	Trauma symptoms
Ewing, MacDonald, Taylor, & Bowers, 2007	Horse training and education	18	Girls with conduct disorders or learning disabilities	28	None	Self-perception Sense of competence Empathy Locus of control Depression Qualitative descriptors
Funahashi et al., 2014	AAT with dog	4	Autism Spectrum Disorder (ASD)	1	Same age peer	Smiles

Table 1: Summary of Included Studies Examining Animal Assisted Therapy for Youth

Citation	Type of AAT	# of Sessions	Sample	Sample Size	Control Group	Dependent Variable
Gabriels et al., 2012	Therapeutic horseback riding	10	Children and adolescents with ASD	42	Wait-list control	Self-regulation Adaptive skills Motor skills
Giagazoglou, Arabatzi, Dipla, Liga, & Kellis, 2012	Hippotherapy	20	Adolescents with intellectual disability	19	No treatment control	Balance Strength
Grigore & Rusu, 2014	AAT with dog	6	Children with ASD	3	None	Appropriate social interaction Need for prompting of responses Initiation of social interactions
Griffioen & Enders-Slegers, 2014	Dolphin-assisted therapy	18	Down syndrome	45	Waitlist/ Crossover	Social skills
Hamama et al., 2011	Dog training	12	Teen girls with trauma history	18	Repeated measures	Subjective well being Coping skills PTSD symptoms
Hanselman, 2001	AAT with dog	12	Teens with anger management problems	7	None	Anger Depression
Hauge, Kvalem, Berget, Ender-Slegers, & Braadstad, 2013	AAT with horse	16	Teens	Study 1: 75 Study 2: 49	Wait-list control	Social support Self-esteem Self-efficacy Mastering horse-related tasks
Heimlich, 2001	AAT with dog	16	Intellectually disabled youth (children and adolescents) in residential facility	14	None	Psychological symptom checklist
Kaiser, Spence, Lavergne, & Vanden Bosch, 2004	Therapeutic horseback riding	5	Able-bodied youth	16	None	Anger Quality of life Perceived self-competence
Keino et al., 2009	Psych-educational horseback riding	Variable	Pervasively developmentally delayed youth	4	None	Human relationships Imitation Emotional expression Sudden physical movement Fixated behavior Adaptation to change Visual response Fear/nervousness Verbal communication Nonverbal communication

Table 1: Summary of Included Studies Examining Animal Assisted Therapy for Youth

Citation	Type of AAT	# of Sessions	Sample	Sample Size	Control Group	Dependent Variable
Kemp, Signal, Botros, Fitchett, Helmer, & Young 1999	Equine-facilitated therapy	9-10	Children and adolescents with history of sexual abuse	30	None	Depression Anxiety Trauma symptoms Internalizing behaviors Externalizing behaviors Problem behaviors
Kogan, Granger, Fitchett, Helmer, Young, 1999	Dog training	11-14	Children with emotional disturbance	2	None	Attention Hyperactivity Social skills Oppositional behavior Vocal expression Eye contact
Kohn & Oerter, 2013	Dolphin-facilitated therapy	Variable	Youth with medical diagnosis	162	None	Cognitive development Social development Language development Emotional development Motor development Conspicuous behavior Observed anger
Lange, Cox, Bernert, & Jenkins, 2007	AAT with dog	Unknown	Teens with anger management problems	5	None	Observed anger
le Roux, Swartz, & Swart, 2014	AAT with dog	10	Children with poor reading skills	102	Read to adult, read to teddy bear, and no treatment	Reading rate Reading accuracy Reading comprehension
Limond, 1997	AAT with dog	7	Children with Down's syndrome	8	Crossover	Looking behavior Responding to the adult Initiating behavior
MacDonald & Cappo, 2003	Equine-facilitated therapy	14	Adolescents in residential treatment	7	None	Perceived competence Locus of control Aggression Self-esteem
Martin & Farnum, 2002	Play therapy with dog	45	Children with PDD	10	Crossover	Prosocial behaviors
Nathanson, 1998	Dolphin-assisted therapy	9-17	Children with physical or medical disabilities	139	None	Adaptive behaviors Targeted behaviors of therapy Parental rating of value of therapy
Nathanson, 1997	Dolphin assisted therapy	17	Children with severe disabilities	47	Physical and speech-language therapy	Independent motor Speech/language skills
Pendry, 2014	Equine-facilitated learning	11	Children with low social competence	113	Waitlist	Social competence Observed positive behavior Observed negative behavior

Table 1: Summary of Included Studies Examining Animal Assisted Therapy for Youth

Citation	Type of AAT	# of Sessions	Sample	Sample Size	Control Group	Dependent Variable
Redefer & Goodman, 1989	Play therapy with dog	18	Children with low IQ and ASD features	12	Repeated measures	Isolation Social interaction
Sams, Fortney & Willenbring, 2006	Occupational therapy with variety of animals	2-12	Children with autism	22	Crossover	Use of language Social interaction
Schuck, Emmerson, Fine & Lakes, 2013	Canine assisted therapy with cognitive behavior therapy plus behavioral parent training	24	ADHD combined type	24	Toy dogs	ADHD symptoms Social skills Social competencies
Schultz, Remick-Barlow & Robbins, 2006	Equine-assisted psychotherapy	1-116 (M = 19)	Children with behavioral and mental health problems	63	None	Global assessment of functioning (GAF)
Signal, Taylor, Botros, Prentice & Lazarus, 2013	Equine-facilitated therapy	9-10	Youth with history of sexual abuse	30	None	Depressive symptoms
Silva, Correia, Lima, Magalhaes & de Sousa, 2011	Canine-assisted therapy	Unknown	Autism Spectrum Disorder	1	Crossover	Observed positive behavior Observed negative behavior
Stumpf & Breitenbach, 2014	Dolphin-assisted therapy	10	Children with severe disabilities	47	No treatment	Communication ability Social emotional behavior Parental quality of life
Taylor et al., 2009	Hippotherapy	16	Autism	3	None	Observed motivation
Thompson, Ketcham and Hall, 2014	Hippotherapy	12	Children with developmental delays or motor deficits	8	None	Posture Walking speed Forward reach Quality of life
Winchester, Kendall, Peters, Sears & Winkley, 2002	Hippotherapy	7	Severe physical disability	7	Repeated measures	Gross motor function Gait speed
Yorke et al., 2012	Equine-assisted therapy	6	Children with PTSD	4	Repeated measures	Cortisol levels
Zasloff, Hart & Weiss, 2003	Humane education and dog training	15	Inner city youth	83	No description	Animal care Fear of dogs Sense of belonging Ability to work in school

Note. Repeated measures control groups indicate studies that utilize multiple baselines and/or follow up designs within each study participant. Crossover control groups indicate within subjects designs in which all participants receive a sequence of different treatment types, alternating over the course of the study (e.g. ABAB)

Quality Coding

Each study included in the review was coded for scientific rigor. The coding system was developed by the first, third, and fourth authors based upon the standards of quality intervention research outlined by Bourton, Moher, Altman, Schulz, and Ravaud (2008), Kendall, Holmbeck, and Verduin (2004), and Kazdin (2010). These guidelines were selected because they address behavioral intervention standards and Consolidated Standards of Reporting Trials (CONSORT) standards for randomized controlled trials, the latter of which are considered to be the gold standard in intervention research. The quality ratings were made in four areas: design, procedure, assessment, and outcome analysis.

To evaluate the design of each study, raters examined the 1) theoretical basis provided for the study, 2) use of a control or comparison condition, and 3) use of random assignment to condition and multiple assessments. To evaluate the procedure of each study, raters examined 4) the description of the sample, 5) description of the intervention, and 6) assessment of treatment fidelity. The rigor of assessments was evaluated in terms of 7) multiple methods of assessment and 8) the use of reliable measures. Finally, the outcome analysis measurement was evaluated in terms of 9) appropriate analytical approaches and 10) testing mediator and moderators of treatment outcomes. Thus, 10 characteristics were rated in each study.

The coders were two advanced clinical psychology doctoral students. Both raters coded twenty-five percent of the studies to assess inter-rater agreement. Studies were coded on a 0 – 2 scale in each category. (See Table 2 for the scoring criteria for each category.) In one case, there was a significant discrepancy (e.g., one rater gave a code of 0 and the other gave a code of 2), between

coders. The coders discussed the specific aspect of the study in question with the third author and a rule was agreed upon to code the characteristic. Coding discrepancy of treatment description was also handled in this way; see further discussion below. The average inter-rater reliability was excellent (weighted $kappa = .97$).

Results and Discussion

The coded data were analyzed by category using SPSS version 22.0 to calculate means, standard deviations, and percent of studies coded in each quality level, and the correlations of the categories with the year of publication. The results suggest that researchers are adhering to a number of optimal principles of empirically supported treatments. The overall quality average of the 45 included studies was 1.04 out of a perfect score of 2.0. The overall quality average was also moderately and positively correlated with the year of publication ($r=.33, p<.05$), which indicates that researchers are applying more rigorous designs in tests of AATs than in the past. However, there was variability across studies, indicating that specific elements of research methodology can be included in the future, to further strengthen the AAT literature and its impact. As described below, researchers can improve overall methodological quality by making several choices regarding the design, procedure, assessment, analysis plan, and conceptual framework, as well as describing their choices clearly in their manuscripts.

Table 2: Quality Coding of Studies

Category	Quality Score		
	0	1	2
Theoretical Basis - Introduction Only	No description	Vague description of theory in the Introduction (e.g., cites empirical findings in the Introduction without larger theoretical rational/framework)	Clear explanation of guiding theory or conceptual framework in the Introduction (e.g., citations to theory/conceptual framework, provides hypotheses based on theoretical or conceptual framework)
Control Group	Single Group Design	Examined group difference for comparison a group that controls for passage of time (e.g., waitlist or treatment as usual)	Examined group differences for a theoretically derived control group (controls for active ingredient or hypothesized mechanism)
Random Assignment	No Random Assignment	Randomly random assignment of existing groups	Randomly assigned to one of two (or more) conditions
Follow-Up Assessment	Single Pre- OR Post-Intervention Assessment	Pre- AND Post-Intervention assessment to assess change (but no follow-up)	Pre-, Post-Intervention AND Follow-up assessment (of any length: one week, one month, etc...)
Defining Sample	No description	Vague description of participants (e.g., unexplained convenience sample, some explanation of participants but little information about sample's demographic characteristics)	Clear description of participants included and excluded (e.g., age, gender, race, clinical problem)
Description of Therapy	No description of administration of intervention	Description insufficient for clinical replication (i.e. 1-2 of the following: specifics of child tasks, skills/processes targeted, techniques used by therapist, type of involvement)	Sufficient information for clinical replication (i.e. 3-4 of the following: specifics of child tasks, skills/processes targeted, techniques used by therapist, type of involvement) or an opportunity to access materials
Treatment Fidelity	No Monitoring	Mention fidelity monitored informally or formally (e.g., mention of supervision sessions, fidelity measures developed for the study; but no statistics reported)	Fidelity monitored formally and results reported (e.g., fidelity results reported, description of efforts taken to ensure adherence)
Reliable Measures	No Reliability Statistics reported for the current sample	Reliability statistics suggest weak/problematic measures (i.e. Cronbach's alpha <.7, or kappa <.8) OR report good reliability statistics from another study but do not report reliability for current sample.	Reliability statistics suggest strong measures in current sample (i.e. Cronbach's alpha in current sample >.7, or kappa >.8)
Multi-Method Assessment of Outcome	No formal assessments (impressions and anecdotes only)	One method of assessment used (e.g., solely self-report, solely informant report)	Multiple methods of assessment (e.g., participant self-report and informant/observer report; parent report and behavioral observation of RA)
Analytic Approach	No statistical significance tests or effect sizes	At least one statistical test or effect size of treatment outcomes reported but not the most appropriate test/statistic for the research question	Appropriate analytic strategies for study design and aims
Mediator	No discussion or testing of mediators	Proposed mediator of treatment anywhere without testing OR Incorrectly tested/interpreted (Mediator is a variable that explains the effect of treatment)	Correctly tested mediators of treatment
Moderator	No discussion or testing of moderators	Proposed moderators of treatment anywhere without testing OR Incorrectly tested/interpreted (Moderator is a variable that explains "for whom" something works and is tested through a multiplicative interaction term)	Correctly tested moderators of treatment

Table 3: Quality Characteristics' Mean Ratings, Distributions Across Studies, And Correlations by Year of Publication.

Quality Characteristic	Quality Rating Mean (SD)	Percent by Rating Score			Correlation of Quality with Year of Publication	Rank
		0	1	2		
Grand Mean Quality Code	1.04 (0.23)				0.33*	8
Theoretical Basis in Introduction	1.76 (0.43)	0	23.91	76.08	0.26	3
Control Group	0.65 (0.82)	56.52	21.73	21.73	0.48**	10
Random Assignment	0.26 (0.68)	86.96	0	13.04	0.30*	12
Follow-Up Assessment	1.15 (0.63)	13.04	58.7	28.26	0.13	7
Defining Sample	1.91 (0.28)	0	8.69	91.3	0.06	1
Duration, Number of Sessions	1.80 (0.48)	4.35	8.69	86.96	-0.12	2
Description of Treatment	1.39 (0.65)	8.69	43.47	47.83	-0.14	5
Treatment Fidelity	0.02 (0.15)	97.83	2.17	0	0.11	14
Reliable Measures	0.98 (0.77)	30.43	41.3	28.26	-0.02	9
Multi-Method Assessment of Outcome	1.22 (0.47)	2.22	73.91	23.91	0.13	6
Analytic Approach	1.67 (0.70)	13.04	6.52	80.43	0.09	4
Mediator	0.50 (0.55)	52.17	45.65	2.17	0.24	11
Moderator	0.22 (0.51)	82.61	13.04	4.35	0.12	13

** p<.01; *p<.05

Rank = Characteristics' quality rankings based on mean ratings with 1 being the best quality

Design

As shown in Table 3, researchers consistently provided theoretical foundations for the use of AATs in the introduction of the papers ($M = 1.76$). Of the studies reviewed, 76.1% provided clear explanations of a guiding theory or conceptual framework in the introduction of the paper. Having a theoretical foundation for the research was not significantly correlated with the year of publication ($r = .26, p > .05$). Theoretical frameworks provide a solid conceptual basis for using and developing AATs. Guiding frameworks also help educate researchers, clinicians, and policy makers with little AAT experience about the value of incorporating AAT in their work. Some of the studies reviewed in this manuscript that included conceptual models focused on attachment theory (Bachi, Terkel, & Teichman, 2014; Balluerka, Muela, Amiano, & Caldenteu, 2014), social learning theory (Schuck, Emmerson, Fine, & Lakes, 2013), Rogerian principals of unconditional positive regard (Kemp, Signal, Botros, Taylor, & Prentice, 2014), or theories of musculature development (Bertoti, 1988; Giagaoglou et al., 2012; Thompson, Ketcham, & Hall, 2014). Continued integration of existing models in the social sciences and other relevant fields in AAT research will help propel the field forward and provide a stronger rationale for the utilization of AATs for specific populations.

In contrast to theoretical foundations, the inclusion of comparison groups is not as common in studies of AATs for youth and has been noted as a significant threat to construct validity in AAT research in general (Marino, 2012). The majority of studies did not include a control or comparison condition (56.5%), whereas 21.7% of studies used a comparison that solely accounted for the passage of time (i.e., waitlist or treatment as usual). Control groups allow researchers to test the extent to which a target intervention is truly efficacious,

while accounting for competing hypotheses such as the passage of time or the attention youth receive in treatment. It is critical that AAT research studies include comparison groups for scientific purposes, but also to demonstrate the value of AATs to clinicians, consumers, and policy makers. The fact that another 21.7% of the studies utilized strong control groups to account for key components of the treatment approach (e.g., unstructured time with an animal, educational components without animal contact) is a good sign, but more studies are needed that include this methodological component. One particularly strong control condition is demonstrated by Dietz, Davis, and Pennings' (2012) use of a dismantling strategy to create two control groups that accounted for two separate key components of the treatment (i.e., therapeutic stories and therapy dogs). Some studies utilized within-subject designs (i.e., crossover or repeated measure designs), which are a useful technique in clinical settings. However, those designs are less desirable for examining efficacy of specific interventions that are expected to have persistent effects, like AATs.

Random assignment is another aspect of design that was generally lacking in the AAT for youth research. Randomly assigning eligible participants to groups minimizes the chance that systematic differences between treatment and control groups can account for group differences. By using random assignment, characteristics such as age, gender, and experience working with animals can be ruled out as alternative explanations of any observed effects. Without random assignment, treatment effects may be attributed to pre-existing differences between groups. Only 13% of studies reviewed used random assignment to place participants in treatment conditions. It should be noted that studies without a control group could not earn points in this category; therefore, of the 20 studies that had comparison groups, six (30%) used random assignment. The other studies used

strategies for group assignment such as convenience sampling, predetermined groups, and healthy controls.

For researchers who have limited budgets or other limiting factors, there are alternatives to random assignment, including matching participants on key variables and/or statistically comparing groups on key variables. Although not ideal, this strategy does control for group differences on variables identified as particularly important in the intervention (e.g., gender, pre-treatment symptoms previous experience with animals). Strength of control conditions was moderately positively correlated with year of publication ($r=.48$, $p<.01$) and the use of random assignment was also moderately positively correlated with year of publication ($r=.30$, $p<.05$). Overall, as interest in this field has grown over the past several years, research methodology in examining efficacy of AATs has become more rigorous by strengthening control conditions and use of random assignment of participants to condition. However, more widespread use of these strong methods is needed.

Pre- and post-treatment assessments were conducted in 58.7% of the studies; however, only 28.3% of studies conducted follow-up assessments. The use of follow-up assessments has not increased as a function of time ($r=.13$, $p>.05$). It is understandable that follow-up assessments can be difficult due to practical barriers, including tracking participants and poor retention rates after treatment, as well as additional costs and resources. However, follow-up provides important information about the lasting effects of AATs beyond the end of treatment. The burden of follow-up studies can be diminished by assessing only key variables at follow-up to lessen the time strain on participants, reducing the interval of the follow-up (e.g., from one month to two weeks), and conducting phone or mail assessments. Stumpf and Breithenbach (2014) provided a good model of follow-up assessment by including a pre-test, a short term

post-test (four weeks), and a longer term post-test (six months).

Overall, the rigor of designs was variable, depending upon the aspects of designs that were evaluated. Researchers consistently provided conceptual frameworks for AATs; however, considerably fewer studies included control groups, random assignment, and repeated assessments including follow-up assessments. Nevertheless, the significant correlations between the year of publication and inclusion of a control group and random assignment are positive indicators that more recent research studies are addressing some design flaws.

Procedure

Clear descriptions of procedures are essential for clinicians and researchers to be able to replicate studies, an important step in demonstrating that an intervention does not work solely in a single setting. Aspects of the procedure that were evaluated included description of the sample, intervention, and treatment fidelity.

Researchers consistently provided information about their study's sample ($M=1.91$); 91.3% of studies reviewed gave a clear description including age, gender, and clinical problems of their participants. A description of the sample allows consumers of AAT research to better interpret the study and determine the applicability of a particular AAT study to their own population of interest. For instance, Braun (2009) made good use of tables to report key descriptors of the sample (i.e., gender, age, pet at home, previous experience with AAT). Our ratings of the duration and number of sessions across these studies also demonstrate that AAT research has adequately described these features ($M=1.83$). Of the studies reviewed, 87% provided a clear description of the length and number of sessions, and an additional 8.7% of studies gave some description, often

describing the number of session participants received but excluding information about the length of sessions. Information about duration and number of sessions serves as a measure of dosage of treatment and allows readers to evaluate the amount of treatment required to obtain desired outcomes as well as determine the utility of a specific AAT in a particular setting. This can often be done in a single sentence in the procedures section.

The description of the intervention is a key to replicability of a study. However, it can be challenging to determine the level of information required for clinical replicability. It was difficult for the current authors to agree on the level of detail needed in this area for it to be considered adequate, proving this to be a challenge in establishing inter-rater reliability in coding how well an intervention was described. Several published guidelines were referenced to create a code that accurately reflected the current standards of the field, including the CONSORT Transparent Reporting of Trials criteria (Bourton, Moher, Altman, Schulz, & Ravaud, 2008) and American Psychological Association's Journal Article Reporting Standards (JARS) (APA, 2008). In the final review process, studies were rated on the clear description of four key characteristics of AAT interventions [i.e., 1) tasks of the youth, 2) specific skills or processes targeted by the intervention, 3) techniques used by therapist to teach the child, and 4) specific type of involvement with the animal] or on whether a study provided a statement about how the details of the intervention could be accessed (e.g., through a specific manual, website, or contacting the author directly). Most studies provided some level of detail regarding the intervention administered ($M=1.39$) and 47.8% of studies provided either enough detail for clinical replication or access to more details. Including details necessary for replication can be a challenge in publications due to manuscript length restrictions. AAT developers should

consider copywriting manuals and making manuals available free online, publishing them, or providing them at the request of interested readers. At a minimum, researchers can indicate that details of their intervention are available by contacting the corresponding author.

Finally, treatment fidelity was monitored informally (e.g., supervision during the intervention) by just 2.2% of studies. None of the studies reported monitoring treatment fidelity in a standardized fashion. Standardized fidelity monitoring could include in-vivo ratings by observers on treatment compliance or participant ratings on treatment at the end of each session. Monitoring the consistent application of AATs within a study provides evidence that each youth participant received the same quality of treatment. Without detailed descriptions of treatment procedures, it is nearly impossible to monitor fidelity; therefore, as the field continues to improve standardization and descriptions of AATs, treatment fidelity will be more easily monitored. Monitoring treatment fidelity has been a universal challenge in behavioral intervention research and continues to be an area that intervention researchers seek to improve (Gearing et al., 2011; Miller & Rollnick, 2014).

Taken together, AAT researchers described components of the intervention procedure to varying degrees. None of the procedural components were correlated with year of publication. Description of the sample and dosage of treatment were clearly stated in the vast majority of studies reviewed. Providing a description of the intervention in detail sufficient for clinical replication was included in nearly half of the studies. However, fidelity of treatment is an area that is significantly lacking in currently published AAT research. As noted above, there are a number of practical steps researchers could model after other intervention research to

improve their descriptions of interventions and treatment fidelity.

Assessment and Measures

Aspects of the assessment and measures that were evaluated included multiple modalities of assessment and reliability of measures. The vast majority of studies used standardized assessments to measure outcomes; however, there is room for improvement in the techniques employed. Of the studies reviewed, 73.9% used one method of assessment (e.g., self-report, observer ratings, or performance on tasks). Only 23.9% of studies included multiple methods of assessing outcomes ($M=1.22$). The use of multiple methods has not significantly changed as a function of the year of publication ($r=.13, p>.05$). The use of multiple methods of assessment provides a comprehensive picture of outcomes and combats the potential bias of a single method (e.g., self-report biases). For instance, Gabriels and colleagues (2012) measured parental self-reported behavioral outcomes using questionnaire and interview data as well as standardized behavioral proficiency tests of motor skills, sensory process, and praxis (the ability to organize, plan, and perform an action). This comprehensive assessment method provides convincing evidence to readers who are not familiar with AATs for youth that these AATs have benefits and should be adopted more widely.

The use of reliable measures for treatment outcomes had an average rating of 0.98 and was not significantly correlated with year of publication ($r=-.02, p>.05$). A total of 28.3% of studies reported strong reliability statistics for their own subjects and 41.3% of studies reported strong reliability statistics based on another sample or weak reliability for the current sample. However, 30.4% of studies did not report the reliability of the outcome measures at all. The results of a study are only

as good as the measures used; therefore, it is essential to use reliable measures to assess treatment outcomes and report the reliability estimates to confirm that the measures have this essential component of validity for the participants who were included. A particularly strong example of reporting reliability statistics is Hamama and colleagues (2011) in which Cronbach's *alpha* coefficients for each outcome measure were reported at pre-test and post-test for the sample.

The selection of measures is often a key consideration in the development of treatment studies. Reporting reliability statistics for the participants of a study demonstrates the use of strong measures to the consumer and builds confidence in the results. Finding appropriate measures for use in AAT research may be difficult due to the relative nascence of the field. In cases where researchers are interested in a specific variable with no previously validated measures, it can be important to design and test measures of the key outcomes before using them to examine the effects of an AAT. This process is often time consuming, but significantly strengthens the findings of AATs. Furthermore, validated measures are more likely to be useful to other AAT researchers, thus benefitting the continued progress of the field.

Outcome Analysis

Sound data analysis plans are essential to the evaluations of AATs for youth. Aspects of each study's outcome analysis that were evaluated included statistical approach, mediator analyses, and moderator analyses. Most researchers appropriately used at least one statistical strategy ($M=1.67$). However, there was no significant correlation with year of publication ($r=.09, p>.05$) indicating that better analytic procedures were not increasing. Only 13% of studies did not use statistical tests to make conclusions about treatment outcomes. The use of statistical tests in examining AAT

treatment outcomes further strengthens the support for conclusions made by the study authors regarding the treatment efficacy. However, mediators or moderators of treatment were tested much less frequently, in only 2.2% and 4.3% of studies respectively; these were not significantly correlated to publication year ($r=.24, p>.05, r=.12, p>.05$). As the research base continues to grow, identifying mediators and moderators will allow researchers and consumers to better understand the mechanisms of treatment and who benefits most of various types of AATs. It should be noted in that mediator and moderator analyses require more statistical power than group differences analyses; therefore, studies with a small samples size lack the ability to test for these variables. One exemplar study in terms of the examination of moderation is le Roux and colleagues (2014) who demonstrated gender moderated the effects of an animal-assisted reading program such that the intervention reduced the existing gender disparity in reading ability compared to control groups. For those researchers not accustomed to using statistical analysis, collaborations with professionals, including qualified graduate students, who have intervention specific data analysis training, could enhance reporting of results through the use of statistical methods.

Conclusions

AAT researchers and consumers have called for more rigorous scientific tests of AATs (Krueger & Serpell, 2010; Kazdin, 2010; Nimer & Lundahl, 2007; Herzog, 2011). This review of AAT for youth demonstrates that the quality of AAT research is variable and that some methodological components have improved over the course of time (e.g., inclusion of a control group and use of random assignment to group). Researchers have much to gain from methodological improvements including greater adoption of AATs by consumers and policy makers, funding for

program evaluation and expansion to reach more youth, and confidence in the results of efficacy and effectiveness trials. Yet other methodological components have not enjoyed greater use over time (e.g., description of treatment, use of reliable measures, mediator and moderator analyses). Thus, the field can continue to improve by addressing methodological weaknesses. It is recommended that AAT researchers make efforts to include strong control groups and random assignment to build the case that AATs for youth can result in positive outcomes that are not due to characteristics of the sample or to other effects such as the passage of time or the mere presence of animal. It is also recommended that researchers provide more detailed treatment descriptions or make their manuals more accessible so that other research groups can attempt to replicate the findings, which would increase the confidence that readers would have regarding the efficacy of the AAT outside the study. Similarly, descriptions allow clinicians to apply scientifically sound practices in the community. Also, monitoring treatment fidelity helps researchers and clinicians alike understand why an AAT may or may not have worked as intended. Demonstrating good treatment fidelity also contributes to internal validity, or the confidence that effects were due to the treatment and not to some other aspect of the intervention (e.g., likeability of therapist). Finally, as researchers build the foundation of support for the effectiveness of AATs, further examination of how it works (mediation) and for whom it works (moderators) becomes increasingly important from both a theoretical perspective and from a practical perspective. As the field grows, future reviews should examine AATs that primarily address physical outcomes and psychosocial outcomes separately to better understand any differences in these treatment domains. AATs for youth show great promise but the ability of the field to convey the importance of this work is

hampered by methodological limitations. It is hoped that this methodological critique can continue to foster the impressive work being undertaken by researchers to demonstrate the efficacy and acceptability of these interventions to the community.

References

Note: Articles with an asterisk (*) were included in the review.

- American Psychological Association (APA). (2008). Reporting standards for research in psychology: why do we need them? What might they be? *American Psychologist*, 63(9), 839-851.
- *Anderson, K. L., & Olson, M. R. (2006). The value of a dog in a classroom of children with severe emotional disorders. *Anthrozoös*, 19(1), 35-49.
- Antonioli, C., & Reveley, M. (2005). Randomised controlled trial of animal facilitated therapy with dolphins in the treatment of depression. *British Medical Journal*, 331(7527), 1231-1234.
- *Bachi, K., Terkel, J., & Teichman, M. (2012). Equine-facilitated psychotherapy for at-risk adolescents: The influence on self-image, self-control and trust. *Clinical Child Psychology and Psychiatry*, 17(2), 298-312.
- *Balluerka, N., Muela, A., Amiano, N., & Caldentey, M. A. (2014). Influence of animal-assisted therapy (AAT) on the attachment representations of youth in residential care. *Children and Youth Services Review*, 42(0), 103-109.
- *Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. *Journal of Autism and Developmental Disorders*, 39(9), 1261-1267.
- *Bertoti, D. B. (1988). Effect of therapeutic horseback riding on posture in children with cerebral palsy. *Physical Therapy*, 68, 1505-1512.
- Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. (2008). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, 148(4), 295-309.
- Bowlby, J. (1982). Attachment and loss: Retrospect and prospect. *American Journal of Orthopsychiatry*, 52, 664-678.
- *Boyer, V. E., & Mundschenk, N. A. (2014). Using animal-assisted therapy to facilitate social communication: A pilot study. *Canadian Journal of Speech-Language Pathology*, 38(1), 26-38.
- *Braun, C., Stangler, T., Narveson, J., & Pettingell, S. (2009). Animal-assisted therapy as a pain relief intervention for children. *Complementary Therapies in Clinical Practice*, 15(2), 105-109.
- *Breitenbach, E., Stumpf, E., Fersen, L. V., & Ebert, H. (2009). Dolphin-assisted therapy: Changes in interaction and communication between children with severe disabilities and their caregivers. *Anthrozoös*, 22(3), 277-289.
- *Cuypers, K., De Ridder, K., & Strandheim, A. (2011). The effect of therapeutic horseback riding on 5 children with attention deficit hyperactivity disorder: A pilot study. *The Journal of Alternative and Complementary Medicine*, 17(10), 901-908.
- *Dietz, T. J., Davis, D., & Pennings, J. (2012). Evaluating animal-assisted therapy in group treatment for child sexual abuse. *Journal of Child Sexual Abuse*, 21(6), 665-683.
- *Ewing, C. A., MacDonald, P. M., Taylor, M., & Bowers, M. J. (2007). Equine-facilitated learning for youths with severe emotional disorders: A quantitative and

- qualitative study. *Child and Youth Care Forum*, 36(1), 59-72.
- *Funahashi, A., Gruebler, A., Aoki, T., Kadone, H., & Suzuki, K. (2014). Brief report: The smiles of a child with autism spectrum disorder during an animal-assisted activity may facilitate social positive behaviors—quantitative analysis with smile-detecting interface. *Journal of Autism and Developmental Disorders*, 44(3), 685-693.
- *Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., . . . Mesibov, G. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6(2), 578-588.
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, 31(1), 79-88.
- *Giagazoglou, P., Arabatzi, F., Dipla, K., Liga, M., & Kellis, E. (2012). Effect of a hippotherapy intervention program on static balance and strength in adolescents with intellectual disabilities. *Research in Developmental Disabilities*, 33(6), 2265-2270.
- *Griffioen, R. E., & Enders-Slegers, M. J. (2014). The effect of dolphin-assisted therapy on the cognitive and social development of children with down syndrome. *Anthrozoös*, 27(4), 569-580.
- *Grigore, A. A., & Rusu, A. S. (2014). Interaction with a therapy dog enhances the effects of social story method in autistic children. *Society & Animals*, 22(3), 241-261.
- *Hamama, L., Hamama-Raz, Y., Dagan, K., Greenfeld, H., Rubinstein, C., & Ben-Ezra, M. (2011). A preliminary study of group intervention along with basic canine training among traumatized teenagers: A 3-month longitudinal study. *Children and Youth Services Review*, 33(10), 1975-1980.
- *Hanselman, J. L. (2001). Animal therapy for coping skills adolescents. *Journal of Child and Adolescent Group Therapy* 11(4), 159-195.
- *Hauge, H., Kvalem, I. L., Berget, B., Enders-Slegers, M.-J., & Braastad, B. O. (2013). Equine-assisted activities and the impact on perceived social support, self-esteem and self-efficacy among adolescents – an intervention study. *International Journal of Adolescence and Youth*, 19(1), 1-21.
- *Heimlich, K. (2001). Animal assisted therapy and the severely disabled child- a quantitative study. *The Journal of Rehabilitation*, 67(4), 48-54.
- Herzog, H. (2011). The impact of pets on human health and psychological well-being fact, fiction, or hypothesis?. *Current Directions in Psychological Science*, 20(4), 236-239.
- *Kaiser, L., Spence, L. J., Lavergne, A. G., & Vanden Bosch, K. L. (2004). Can a week of therapeutic riding make a difference?— A pilot study. *Anthrozoös*, 17(1), 63.
- Kazdin, A. E. (2010). Methodological standards and strategies for establishing the evidence base of animal-assisted therapy. In A. E. Fine (Ed.), *Handbook on Animal-Assisted Therapy: Theoretical Foundations and Guidelines for Practice* (519-546). Burlington, MA: Academic Press.
- Kazdin, A. E., & Weisz, J. R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology* 66(1), 19-36.
- *Keino, H., Funahashi, A., Keino, H., Miwa, C., Hosokawa, M., Hayashi, Y., & Kawakita, K. (2009). Psycho-educational horseback riding to facilitate

- communication ability of children with pervasive developmental disorders. *Journal of Equine Science* 20(4), 79-88.
- *Kemp, K., Signal, T., Botros, H., Taylor, N., & Prentice, K. (2014). Equine facilitated therapy with children and adolescents who have been sexually abused: A program evaluation study. *Journal of Child and Family Studies*, 23(3), 558-566.
- Kendall, P. C., Holmbeck, G., & Verduin, T. (2004). Methodology, design, and evaluation in psychotherapy research. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed., pp. 16-43). New York: Wiley.
- *Kogan, L., Granger, B., Fitchett, J., Helmer, K., & Young, K. (1999). The human-animal team approach for children with emotional disorders: Two case studies. *Child and Youth Care Forum*, 28(2), 105-121.
- *Kohn, N., & Oerter, R. (2013). Dolphin assisted therapy works: scientific findings from Eilat and Florida. *International Journal of Clinical Psychiatry*, 1(1), 1-16.
- Kruger, K. A., & Serpell, J. A. (2010). Animal-assisted interventions in mental health: definitions and theoretical foundations. In A.H. Fine (Ed.). *Handbook on Animal-Assisted Therapy: Theoretical Foundations and Guidelines for Practice* (33-48). Burlington, MA: Academic Press.
- *Lange, A. M., Cox, J. A., Bernert, D. J., & Jenkins, C. D. (2007). Is counseling going to the dogs? An exploratory study related to the inclusion of an animal in group counseling with adolescents. *Journal of Creativity in Mental Health*, 2(2), 17-31.
- *le Roux, M. C., Swartz, L., & Swart, E. (2014). The effect of an animal-assisted reading program on the reading rate, accuracy and comprehension of grade 3 students: A randomized control study. *Child & Youth Care Forum*, 43(6), 655-673.
- *Limond, J. A., Bradshaw, John W.S., Magnus Cormack, K.F. (1997). Behavior of children with learning disabilities interacting with a therapy dog. *Anthrozoös: A Multidisciplinary Journal of The Interactions of People & Animals*, 10(2/3), 84-89.
- *MacDonald, P., & Cappo, J. (2003). Equine Therapy with at-risk youth: Does it work? *Strides*, 9(3), 30-31.
- Marino, L. (2012). Construct validity of animal assisted therapy and activities: How important is the animal in AAT? *Anthrozoös: A Multidisciplinary Journal of The Interactions of People & Animals*, 25(1), 139-151.
- * Martin, F., & Farnum, J. (2002). Animal-assisted therapy for children with pervasive developmental disorders. *Western Journal of Nursing Research*, 24(6), 657-670.
- Maujean, A., Pepping, C., & Kendall, E. (2015). A systematic review of randomized controlled trials of animal-assisted therapy on psychosocial outcomes. *Anthrozoös*, 28(1), 23-36.
- Miller, W. R., & Rollnick, S. (2014). The effectiveness and ineffectiveness of complex behavioral interventions: impact of treatment fidelity. *Contemporary Clinical Trials*, 37(2), 234-241.
- *Nathanson, D. E. (1998). Long-term effectiveness of dolphin-assisted therapy for children with severe disabilities. *Anthrozoös*, 11(1), 22-32.
- *Nathanson, D. E., de Castro, D., Friend, H., & McMahon, M. (1997). Effectiveness of short-term dolphin-assisted therapy for children with severe disabilities. *Anthrozoös*, 10(2-3), 90-100.
- Nimer, J., & Lundahl, B. (2007). Animal-assisted therapy: A meta-analysis. *Anthrozoös*, 20(3), 225-238.

- Pet Partners (2012). About animal-assisted activities and animal-assisted therapy. Retrieved from: <http://www.petpartners.org/page.aspx?pid=320>.
- *Pendry, P., Carr, A. M., Smith, A. N., & Roeter, S. M. (2014). Improving adolescent social competence and behavior: A randomized trial of an 11-week equine facilitated learning prevention program. *The Journal of Primary Prevention, 35*(4), 281-293.
- *Redefer, L.A. & Goodman, J.F. (1989). Brief report: Pet-facilitated therapy with autistic children. *Journal of Autism and Developmental Disorders, 19*(3), 461-467.
- *Sams, M. J., Fortney, E. V., & Willenbring, S. (2006). Occupational therapy incorporating animals for children with autism - a pilot investigation. *American Journal of Occupational Therapy, 60*(3), 268-274.
- *Schuck, S. E. B., Emmerson, N., Fine, A. H., & Lakes, K. D. (2013). Canine-assisted therapy for children with ADHD: Preliminary findings from the positive assertive cooperative kids study. *Journal of Attention Disorders, (23)*13, 1-13.
- *Schultz, P. N., Remick-Barlow, G. A., & Robbins, L. (2006). Equine-assisted psychotherapy: a mental health promotion/intervention modality for children who have experienced intra-family violence. *Health & Social Care in the Community, 15*(3), 265-271.
- *Signal, T., Taylor, N., Botros, H., Prentice, K., & Lazarus, K. (2013). Whispering to horses childhood sexual abuse, depression and the efficacy of equine facilitated therapy. *Sexual Abuse in Australia and New Zealand, 5*(1), 24-32.
- *Silva, K., Correia, R., Lima, M., Magalhaes, A., & de Sousa, L. (2011). Can dogs prime autistic children for therapy? Evidence from a single case study. *Journal of Alternative and Complimentary Medicin, 17*(7), 655-659.
- Souter, M. A., & Miller, M. D. (2007). Do animal-assisted activities effectively treat depression? A meta-analysis. *Anthrozoos, 20*(2), 167-180.
- *Stumpf, E., & Breitenbach, E. (2014). Dolphin-assisted therapy with parental involvement for children with severe disabilities: Further evidence for a family-centered theory for effectiveness. *Anthrozoös, 27*(1), 95-109.
- *Taylor, R. R., Kielhofner, G., Smith, C., Butler, S., Cahill, S. M., Ciukaj, M. D., & Gehman, M. (2009). Volitional change in children with autism: A single-case design study of the impact of hippotherapy on motivation. *Occupational Therapy in Mental Health, 25*(2), 192-200.
- *Thompson, F., Ketcham, C. J., & Hall, E. E. (2014). Hippotherapy in children with developmental delays: Physical function and psychological benefits. *Advances in Physical Education, 4*(2), 60-69.
- *Winchester, P., Kendall, K., Peters, H., Sears, N., & Winkley, T. (2002). The effect of therapeutic horseback riding on gross motor function and gait speed in children who are developmentally delayed. *Physical & Occupational Therapy in Pediatrics, 22*(3-4), 37-50.
- *Yorke, J., Nugent, W., Strand, E., Bolen, R., New, J., & Davis, C. (2012). Equine-assisted therapy and its impact on cortisol levels of children and horses: a pilot study and meta-analysis. *Early Child Development and Care, 183*(7), 874-894.
- *Zasloff, L. R., Hart, L. A., & Weiss, J. M. (2003). Dog training as a violence prevention tool for at-risk adolescents. *Anthrozoös, 16*(4), 352-359.