Running head: HUMANE EDUCATION

# Improving Upper Elementary Students' Humane Attitudes and Prosocial Behaviors through an In-Class Humane Education Program

William Ellery Samuels*

The City University of New York

College of Staten Island

School of Education

Lieve Lucia Meers

Belgian Institute for Animal Assisted Therapy

Simona Normando

Università di Padova

Dipartimento di Biomedicina Comparata e Alimentazione

This is an Accepted Manuscript of an article published by Taylor & Francis Group in *Anthrozoös*, available online: http://www.tandfonline.com/doi/abs/10.1080/08927936.2016.1228751.

## Abstract

Humane education is a long-standing field of education that endeavors to nurture kindness, compassion, and concern for non-human animals, people, and the environment. Despite its long history, however, few randomized control studies have evaluated its effectiveness to promote the development of relevant outcomes. The current study sought to address this dearth by investigating the effects of a humane education program on not only participating students' attitudes, but also behaviors. Classes of under-served, fourth-grade students in in two major U.S. cities were randomly assigned to participate in either an school-based humane education program or a chess club (the control group); all students participated in their respective activity during the same period once a week for 11 weeks. Data were collected on the attitudes ($N$ = 236) and behaviors ($N$ = 167) of participating students exactly one week before and exactly one week after these programs were conducted. Students self-reported attitudes about the treatment of animals and the environment via the Intermediate Attitude Scale. Teachers rated each students' prosocial and disruptive behaviors though respective sub-scales of the Teacher Observation of Classroom Adaptation–Checklist. Nested multilevel models of change found that the development of prosocial behaviors and self-reported attitudes significantly interacted with group assignment: Students who participated in the humane education program showed stronger growth in both compared to students in the control group. Changes in disruptive behaviors, however, were not found to differ between groups. The results support the effectiveness of a humane education program to teach a relatively large and diverse group of upper elementary students to learn about animal welfare issues and to improve their prosocial behaviors. Effects appeared strongest on attitudes; behavioral effects were

found to be largely limited to behaviors directly addressed by the humane education program.

Key words

Humane education, elementary school, prosocial behavior, disruptive behavior, humane attitudes, program evaluation

# Upper Elementary Students' Humane Attitudes and Prosocial Behaviors Can Be Improved Through an In-Class, Humane Education Program

## *Humane Education: Definition and Prominence*

Although the focus and methods of humane education have changed since its creation as a discrete field in the late 1800s (Preston, 1928), it has always promoted the humane treatment of human and non-human animals (Burnett, 2000). Humane education enjoyed popular support during the late 19th and early 20th centuries, but its prominence in North America and Europe declined in the years following the world wars in part because there was that little scientific evidence proffered in its defense (Unti & DeRosa, 2003). This, however, was not due to evidence indicating that humane education was *not* effective, but simply that very little research was conducted to support it.

Nonetheless, humane education has remained a staple of both local animal rescue and shelter organizations (Faver, 2010). It is also been championed by many locally-active and national organizations which conduct a wide array of humane education programs in schools, at extra-curricular events, summer camps, community centers, etc. Humane education therefore remains widely—if diffusely—practiced, relying largely on humane educators' experience and pedagogical knowledge to fill in the gaps that research has left open.

## *Review of Research on Humane Education's Effectiveness*

The research on humane education has typically included field-based evaluations of changes in children's self-reported attitudes towards animals as well as the environment and other people. The limited research that exists can be roughly grouped by the instruments used to assess the changes in attitudes.

*The Fireman Test*

The first scientific, quantitative studies of humane education tended to use versions of a *Fireman Test*. Developed and tested by Vockell (1979); this test asks students which three of ten items/entities they would choose to have rescued from a burning building; three of the choices were animals, the rest were inanimate objects (TV, stereo system, etc.). Vockell and Hodal (1980) used the Fireman Test to measure the outcomes of a study in which they randomly assigned elementary school students in several schools to one of three groups: those participating in a one-period humane education program and receiving humane-education-related materials on two occasions, those receiving only the humane education materials, and those receiving no intervention (a control group). Vockell and Hodal found that, on one version of the Fireman Test, students who had participated in the humane education program on two occasions chose to save animals more often than students assigned the control group receiving no programming.

Fitzgerald (1981) also found that post-program scores on the Fireman Test were significantly higher (i.e., animals were saved more often) for fifth and sixth grade students who participated in a four-lesson humane education unit plus printed materials than for similar students who participated in a one-lesson program plus materials, who received materials alone, or who received no lessons or materials. They found no differences between these latter three groups. In other words, Fitzgerald found that attitudes towards animals were affected by repeated lessons and not by brief programming or printed materials. A similar study conducted by Malcarne (1983) found some evidence that a humane education program was more effective when conducted over multiple sessions than through one, intensive session.

*The Primary Attitude Scale and the Intermediate Attitude Scale*

The Western Institute for Research & Evaluation and the National Association for Humane & Environmental Education (WIRE & NAHEE, 1983) designed a pair of instruments to measure children's self-reported humane attitudes towards various types of animals (pets, farm animals, and wild animals), other people, and the environment. The first instrument, the *Primary Attitude Scale* (PAS), measures these attitudes in lower elementary students; the second, the *Intermediate Attitude Scale* (IAS), measures them in upper elementary students. The IAS is described in more detail in the Methods section, below. These instruments have enjoyed relatively wide use and have served as the yardstick for most measurements of elementary students' humane attitudes, i.e., their attitudes about issues commonly addressed through humane education.

The first published uses of the PAS and IAS were in an ambitious line of research conducted by Ascione and his colleagues. In the first study, Ascione (1992) administered either the PAS or the IAS (along with other instruments) to students in 16 classrooms that were randomly assigned to either an experimental or control group. Students in the experimental group participated in roughly 40 hours of humane education programming created through NAHEE; students in the control group had no special programming during the study. Ascione found that fourth (but not first, second, or fifth) grade students who had participated in the humane education program demonstrated significant gains in humane attitudes relative to those in the respective control groups. Participating in the program also improved fourth and fifth grade students' self-reported empathy towards people. Following up with the fourth-grade participants during the subsequent academic year, Ascione and Weber (1996) found that the effect persisted, albeit slightly attenuated.

Fonseca et al. (2011) also used the IAS to evaluate a rather large sample of upper elementary school students, half of whom participated in a RODENTIA program in which students learned about the life sciences, scientific inquiry, and animal welfare through course-based activities that employed lab rats (*Rattus norvegicus*) housed in the classrooms during the academic year and cared for in part by the students. Fonseca et al. found that students who participated in the program showed significant increases in humane attitudes at the end of the academic year compared to the beginning of the year; students participating in their non-programming control group showed no significant changes in these attitudes.

Nicoll, Trifone, and Samuels (2008) also found that students' attitudes towards animals, etc. measured by the PAS improved after participating in a six-lesson humane education program compared to students who participated in an equal-length, non-human education program. The researchers did not find an effect of exposing students only to humane education materials for two months.

*Limits of Research on Humane Education and Justification for the Current Study*

These and a few other published studies (e.g., Arbour, Signal, & Taylor, 2009; Aguirre & Orihuela, 2010; and Finch, 1984) report that humane education programs can improve students' knowledge, attitudes, or even self-reported empathy towards others. Although this is an encouraging start, it is only a start. One primary limitation of previous studies on humane education is that they only measured students' anticipated behaviors through self reports.

We attempted to address this issue in the current study by asking students' teachers to report on relevant changes in their students' behaviors. We did this through an in-

strument developed for general use in classrooms to assess students' overall frequency of behaviors pertaining to prosociality and disruptiveness.

We also used the IAS to measure students' humane attitudes. Doing so not only allows a direct assessment of the current humane education program, but also provides an indirect gauge of how much a humane education program affects behaviors compared to attitudes.

*Hypotheses*

The currently-evaluated humane education program employs student-centered approaches similar to those of the programs evaluated by Ascione (1992, Ascione & Weber, 1996); Arbour, Signal, & Taylor (2009); Fonseca et al. (2011); and Nicoll, Trifone, and Samuels (2008). Given the program content as well as the extent to which students are encouraged to work cooperatively throughout the program, we hypothesized that the program will lead to general gains in prosocial behavior compared with the gains made by students in a non-humane education, control program.

The humane education program is not designed to address disruptiveness directly, and we measured general disruptiveness—not just that displayed during the program, a time when students may be less disruptive simply because they are currently engaged. Nonetheless, the respect that the humane education program seeks to nurture may lead students to be more mindful of the effects of negative in-class behaviors on those around them. Therefore, we also hypothesized that students participating in the humane education program may be less disruptive than those participating in the control program.

Finally, we hypothesized that the humane education program will also lead to greater changes in attitudes towards humane issues than does the control program. After

participating in the humane education program, we hypothesized that students will self-report being more aware of and concerned about humane issues than do students after participating in the control program.

## Methods

### *Participants*

A total of 284 fourth-grade students participated in the study, 140 of whom were female and 139 of whom were male, and 5 of whom did not self-report gender. Although we did not measure individual students' ages, all participants were traditional students whose ages ranged from 9 to 11 years. After attenuation from missing data, the data from 236 students were available for analyses with the IAS and 167 were available for analyses with the TOCA-C. Students attended one of four, under-served Title 1 public schools (defined as having at least 40% of the student body eligible for free/reduced-price school lunches), two in Chicago and two in New York City. No ethnicity or socio-economic data were collected from the students. Table 1, however, presents the demographic information for each of the schools' total populations. Students were from 12 different classes, 3 from each school. The minimum class size was 17, the maximum was 45, and the mean was 23.67 students.

### *Missing Data and Protocol Deviations*

#### *Missing Data*

All students and their parents agreed to participate in the study. However, any students not in attendance on a day of administration did not complete those instruments. Because of this, 33 students (24 from the experimental group) were not administered the instruments during the pre-program, and 15 (10 from the experimental group)

were not administered them at post-program. Fisher's exact tests found no significant differences between the proportion of missing data from the experimental and control groups at either pre-program ($p$ = 0.83) or post-program ($p$ = 0.55).

As recommended by Altman et al. (2001), before the study was conducted, we established that participants' data would not be used if 10% or more of an instrument's items were left blank. Insufficiently-completed instruments led to the removal of 2 students from analyses of the TOCA-C; one of whom was from a pre-program experimental group and one from a post-program experimental group. Fisher's exact tests again found no significant difference in the proportions of missing data between the experimental and control groups ($p$s ≈ 1).

Eleven IAS scores were removed due to blank items, 9 of which were from the pre-program. These differences were also not found to be significantly different (pre-program $p$ = 0.72; post-program $p$ = 0.56).

In addition, 35 students did not give sufficient information to allow their responses to be linked with either their class or their other data (e.g., students did not report their first name or their teacher's last name) and therefore could not be used for analyses. In all, varying amounts of data were missing for 79 students.

How one should address missing data depends on why they are missing. We assessed whether these data could be considered missing at random through the Fisher's exact tests described above, which tested the hypothesis that the proportion of missing data were different between the experimental and control groups. We found no evidence that the proportion of missing data differed between the experimental and control groups. We therefore treated these missing data as missing at random. Given this, we

conducted analyses using the cases with data available for the given analysis. Therefore, the sample size for the analyses of TOCA-C subscores differs from that for the analyses of IAS scores.

*Protocol Deviations*

One of the schools, labeled in the tables as School D, was affected by a natural disaster between the end of the programs and the collection of post-program data. Although the students at this school completed the IAS anyway, the teachers at this school did not complete the TOCA-C. These TOCA-C data can be considered missing completely at random, allowing their exclusion not to greatly affect interpretation of the respective results (Allison, 2001). Since there are no post-program TOCA-C data from this school available, 167 students are available for analyses with the TOCA-C. Following an intent-to-treat analysis protocol (Armijo-Olivo, Warren, & Magee, 2009; Unnebrink & Windeler, 2001), the post-program data for IAS scores from this school were treated similarly to the data from the other schools.

## *Materials*

*Circle of Compassion Humane Education Program*

The Circle of Compassion program was conducted once a week for 11 weeks with upper elementary students. A trained humane educator visited the students in their classroom during a regularly-scheduled class period. The program used age-appropriate, multimedia, student-centered activities and service learning events—in addition to some traditional teaching strategies—to familiarize students with typical humane education topics and ways that the students can improve the well-being of others. The program covered challenges faced by pets, farm animals, wildlife, the environment, and children around

the world through discussions and explorations of issues that include animal welfare, blood sports, conservation, pollution and climate change, bullying, diet and consumer choices, poverty, and child labor. The program was created and conducted by Humane Education Advocates Reaching Teachers (HEART), a non-profit organization that serves elementary-school students through materials and programs that "focus on human rights, animal protection and environmental ethics" (HEART, 2014). The educators who conducted the program included HEART employees and unpaid volunteers, all of whom had prior experience teaching elementary-aged students.

*Measurement Instruments*

*Teacher Observation of Classroom Adaptation–Checklist*

The Teacher Observation of Classroom Adaptation–Checklist (TOCA-C; Leaf, Schultz, Keys, & Ialongo, 2002; based on the Teacher Observation of Classroom Adaptation, Kellam, Branch, Agrawal, & Ensminger, 1975) asks teachers or other school-based professionals to inventory how frequently a given student has demonstrated 21 socially adaptive behaviors in the last three weeks. The various behaviors are divided into three sub-scales, two of which are studied here: prosocial behavior (e.g., is friendly, shows empathy & compassion for others' feelings) and disruptiveness (e.g., breaks rules, harms property). The third sub-scale, concentration, was not studied since it is not theoretically relevant.

There is considerable evidence that the TOCA-C validly measures its purported domains (e.g., Schaeffer, Petras, Ialongo, Poduska, & Kellam 2003; Petras, Chilcoat, Leaf, Ialongo, & Kellam, 2004; Schaeffer et al., 2006; Koth, Bradshaw, & Leaf, 2009), and the internal reliability of the sub-scales is strong (Cronbach's alphas from 0.85 – 0.91). Versions

of the TOCA, including the one employed here, have served numerous research and clinical uses.

The TOCA-C asks teachers to rate each student's behavior over the past three weeks in general. Doing so, the TOCA-C measures changes in students' behaviors throughout their time at school in the presence of teachers—not only while actively participating in the experimental or control program. The TOCA-C, therefore, measures not only changes in behaviors, but indirectly whether these changes generalize to occasions outside of the program.

*Intermediate Attitude Scale*

The Intermediate Attitude Scale (IAS; Ascione, 1988) measures upper elementary school children's attitudes about and knowledge of animal welfare (concerning companion, farm, and wild animals) and environmental issues. It is comprised of 31 items that ask respondents to report how much they agree or disagree with various statements (e.g., "keeping farm animals in small spaces is not good even if it increases food production," "none of the needs that animals have are similar to human needs," and "a littered environment is a bad environment for most animals").

Ascione (1992) found the IAS to have a Cronbach's alpha of 0.69; for comparison's sake, the Cronbach's alpha for the pre-program scores in the current study was 0.64, and for the post-program scores it was 0.70. These are relatively weak values for an instruments' Cronbach's alpha. The IAS measures students' responses to several issues and activities related to domestic, feral, and wild animals. Therefore, the IAS may be measure more than one—if still non-orthogonal—dimensions.

*Procedure*

With Institutional Review Board approval (CUNY HRPP #12-01-003-0146; NYC-DOE IRB file #65), we administered the IAS to students in four different schools, all schools in which HEART had previously conducted the Circle of Compassion program with other students in earlier years. Classes were randomly assigned to participate in either the Circle of Compassion humane education program (the experimental group) or a chess club (the control group); both programs were conducted one period per week for 11 weeks. All students in a given class participated in the same program. All students were administered the instruments one week before and exactly week after the programs started. The principal investigator, who is unaffiliated with HEART and received no compensation for the research, supervised students' completion of the instruments in New York City; a HEART representative who did not instruct any of these students supervised the administration of the instruments in Chicago. Students who were absent on either day of administration were excluded from those analyses but still participated in their respective program.

The TOCA-C was given to the students' teachers to complete while the students completed the IAS. The teachers completed the TOCA-C once for each student both before and after the study. Teachers who did not complete the TOCA-C for each student during the class period were allowed one week to complete it.

Exactly one week after the students completed the pre-program data collection, all students participated in an in-class enhanced education program. Those in the experimental group participated in the Circle of Compassion program as it is typically conducted. Those in the control group participated in a chess club, first learning the rules of the game and some basic strategies and then playing and discussing the game during the

same class period and for the same number of weeks that the Circle of Compassion was being conducted. Since three classes were available at each school to participate, we balanced the effect of school and assign two classes in each school (i.e., eight total classes) to the experimental group and one class from each school (i.e., four total classes) to the control group. Data were collected for the post-program exactly one week after students completed their respective program the same way they were collected for the pre-program.

*Analyses*

After assessing whether the data could be treated as normally distributed, we conducted a series of multi-level change models in which time (pre-program and post-program) where included in the level 1 (within-student) sub-model. In the level 2 (between-student) sub-model, students were nested within school and school was nested within city.

A taxonomy of three multi-level change models were run for each dependent variable (i.e., three models each for prosociality, disruptiveness, and IAS scores). The first model in the taxonomy included only a term for time and no terms for the treatment (experimental versus control). The main effect for treatment was added to the second model in the taxonomy. The treatment x time interaction term was added to the third model. By adding one term at a time, we gained an additional test of the significance of the added term: In addition to testing whether that term was significant (i.e., that the term's β-weight was significant), we could also test whether the new model including that term was a significantly better fit to the data. This is roughly analogous to testing the difference in $R^2$ between the two models.

In addition, we tested whether nesting students within school and whether nesting schools within city helped clarify the models by computing intra-class correlations (ICCs) for each level for the final models (i.e., the third model in each of Tables 3, 4 and 5). For students nested within school, the ICCs were 0.30, 0.17, and 0.66 for prosocial, disruptiveness, and IAS scores, respectively. For nesting within city, the ICCs were 0.20, 0.02, and 0.65, respectively. (Higher ICCs indicate that more of the variance in the given score is due to differences between the groups of the nested variable.) The nesting variable accounted for considerable variance in all cases except for disruptiveness score data nested within city where only 2% of the variance in disruptiveness scores was accounted for by which city the students resided. Since nested appeared justified for all models except disruptiveness scores within city, we retained all nesting; this allowed for readier comparisons of models between scores—especially between prosocial and disruptiveness models—while having only a minor effect on the stochastic terms in the disruptiveness models.

All scores were standardized before being added to the models. This allows easier comparisons of β-weights between models; β-weights also represent partial correlations. It also obviates the need for an intercept term in the models; since a score of zero on any of the instruments has no meaning, there was no need for an intercept term.

All analyses were conducted with R, version 3.0.2 (R Core Team, 2013). R packages used included psych (Revelle, 2014) and nlme (Pinheiro et al., 2013).

## Results

### *Descriptive Statistics*

Table 2 presents the descriptive statistics for the TOCA-C prosociality, TOCA-C disruptiveness, and IAS total scores. Figures 1 - 3 present the marginal means and 95%

confidence intervals for the effects of the Circle of Compassion and control group programs over time. Shapiro-Wilks tests of normality (as well as visual scans of the data) did not find that any of the pre-program or post-program scores for either the Experimental or Control groups deviated from normality (smallest $p = 0.072$ for post-program prosociality scores).

*Teacher Ratings of Students' Pre- and Post-Program Prosociality*

Figure 1 depicts the pre- and post-program marginal mean TOCA-C prosocial scores for the experimental and control groups along with the 95% confidence intervals. Table 3 summarizes the results of the multi-level change models testing the effect of the program on students' prosocial behavior. The first model in Table 3 presents the results when only time (i.e., change in scores from pre-program post-program testing) was included; model 2 presents the results when a main treatment effect was added; and model 3 presents the results when a treatment x time interaction term was also added.

In addition to testing whether individual terms are significant, we gain an additional test of significance through analyzing how well each of these three models fit the data. This assessed by comparing whether the difference in −2 log likelihoods (−2*LL*s) between two models is greater than a pre-determined critical value. For the test between adjacent models, the critical $\chi^2_1 = 3.84$; for the difference between model 1 and 3, the critical $\chi^2_2 = 5.99$. We can see, therefore, that adding the treatment x time interaction produced a model that fit the data better than models 1 and 2. Those in the experimental group tended to experience greater growth in prosociality than did those in the control group.

*Disruptiveness*

Figure 2 presents the marginal mean TOCA-C disruptiveness scores for the experimental and control groups both before and after the programs were conducted. Lower scores indicate more disruptive behaviors, so in Figure 2, it appears that students may have been rated as more disruptive after the programs compared with before them, however the error bars suggest this difference was not significant. Similarly, although students who participated in the Caring for Life program were rated as somewhat less disruptive both before and after the program, this difference did not change (the lines appear to be roughly parallel) and, in any case, does not appear to be significant.

The analyses support these impressions. Although adding a main treatment effect term allows the model to fit the data significantly better (difference in $-2LLs = 45.72$), the treatment term itself is not significant. Adding the treatment x time interaction term does not significantly improve the model fit and only adds another non-significant term to the model.

*Students' Self-Reported Attitudes towards Animals and the Environment*

Figure 3 depicts the changes in IAS scores for the two groups over time. This figure shows that the IAS scores for the students in the experimental group strongly increasing. However, the IAS scores for the students in the control group appear to also increase. As the differences in $-2LLs$ between model 3 and either model 1 or 2 in Table 5 show, students in the experimental group demonstrated a greater increase in the rate at which their attitudes developed than students in the control group.

## Discussion

Under-served, upper-elementary students who were randomly assigned to participate in the Circle of Compassion humane education program were rated as becoming

more prosocial and reported having more humane attitudes after the program than students who were assigned to participate in a non-humane education program. The humane education program did not have a significant effect on students' demonstrated disruptiveness, however.

We hypothesized that the frequency of prosocial behaviors would increase and that the frequency of disruptive ones might decrease. However, we only found evidence that prosocial behaviors were affected. Disruptiveness was not significantly affected by the program that the students participated in. Any differences in disruptiveness between the two groups can be ascribed to chance. It therefore appears that the students are learning to actively behave more prosocially, but are not also learning to curtail their disruptive behaviors.

The Caring for Life program directly addresses prosociality through content and activities directed toward both other people as well as animals and the environment. It does not directly address disruptive behaviors, however, we hypothesized that the program's inclusion of respect for others could possibly translate into being concomitantly less disruptive. We did not find support for this conjecture, though; the increased development of prosocial behavior among students in the humane education program did not come with proportional losses in disruptive behaviors. Program creators may therefore need to explicitly include content and activities focused on disruptiveness if they wish to see changes there.

Although the types of behaviors affected by the humane education program were constrained to those explicitly addressed by the program, the changes in these behaviors likely generalized to occasions outside of the program. Although we did not specifically

ask the teachers to rate how the children acted outside of the program (doing so would have required a change in the instrument), the instrument does not ask the respondents to limit their ratings to particular occasions, and informal conversations with most of the teachers indicated that they indeed rated the children on how they behaved in general over the last three weeks—not just in the relatively little time they spent participating in the program.

Students who participated in the humane education program also reported greater awareness of and concern for humane issues than did those who participated in the chess club. The program therefore appears effective at changing upper elementary students' attitudes about these issues. The changes in attitudes appeared somewhat stronger than the changes in behaviors.

Finding changes in knowledge and attitudes among those who participated in the Caring for Life program replicates the effects of humane education programs found by, e.g., Ascione (1992) and Fonseca et al. (2011). In general, it appears that various humane education programs can effectively teach children content and change their attitudes. The current study extends this conclusion by finding that a humane education program can also affect behaviors that are addressed by the program.

It should be noted that the topics addressed in the Circle of Compassion are typical of humane education programs, but the pedagogical strategies used here may differ from other humane education programs. The program utilizes many student-centered and experiential activities; students often work in small groups to use what they learn to plan and sometimes implement strategies to help animals, other children, and the environment. These strategies may have affected the program's success, and future research may

benefit from investigating the role pedagogy plays in effective humane education programming.

In addition, as the overall effectiveness of humane education programs becomes increasingly well established, future studies may also begin to distinguish between which areas of humane education are most able to change and which areas require more prolonged efforts for humane educators and school-based professionals.

*Limitations*

In addition to the self-report nature of the assessments of humane attitudes, the teachers who rated the students' behaviors were not blind to the students' group assignment. The sample sizes of the experimental and control groups were different, which may have reduced our ability to detect an effect of the programming (i.e., increased the chance of a Type II error). Finally, the educators who conducted the Circle of Compassion program were highly trained and very used to conducting this program with similar students; it may therefore be that programs conducted by less experienced educators would realize weaker results.

References

Aguirre, V. and Orihuela, A. 2010. Assessment of the impact of an animal welfare educational course with first grade children in rural schools in the state of Morelos, Mexico. *Early Childhood Education Journal* 38: 27 - 31.

Allison, P. D. 2001. *Missing data.* Sage University Series on Quantitative Application in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gotzsche, P. C., and Lang, T. 2001. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine* 134: 663 – 694.

Armijo-Olivo, S., Warren, S., and Magee, D. 2009. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: A review. *Physical Therapy Reviews* 14, 36 – 49.

Arbour, R., Signal, T., and Taylor, N. 2009. Teaching kindness: The promise of humane education. *Society & Animals* 17, 136 – 148.

Ascione, F. R. 1988. *Intermediate Attitude Scale. Assessment of third through sixth graders' attitudes toward the treatment of animals.* Logan, UT: Wasatch Institute for Research Evaluation.

Ascione, F. R. 1992. Enhancing children's attitudes about the humane treatment of animals: Generalization to human-directed empathy. *Anthrozoös* 5, 176 – 191.

Ascione, F. R. and Weber, C. V. 1996. Children's attitudes about the humane treatment of animals and empathy: One-year follow up of a school-based intervention. *Anthrozoös* 9, 188 - 195.

Burnett, C. 2000. Humane education. *Animals Today* 8: 18 - 20.

Faver, C. A. 2010. School-based humane education as a strategy to prevent violence: Review and recommendations. *Children & Youth Services Review* 32: 365 – 370.

Finch, P. 1984. A look at humane education summer camps. *Humane Education* 8: 18 – 21.

Fitzgerald, T. A. 1981. Evaluating humane education: The Jefferson County study. *Humane Education* 5: 21 – 22.

Fonseca, M. J., Franco, N. H., Brosseron, F., Tavares, F., Olsson, I. A. S., and Borlido-Santos, J. 2011. Children's attitudes towards animals: Evidence from the RODENTIA Project. *Journal of Biological Education* 45: 121 – 128.

Humane Education Advocates Reaching Teachers. 2014. Home page. <http://teachhumane.org/> Accessed February 28, 2016.

Kellam, S. G., Branch, J. D., Agrawal, K. C., and Ensminger, M. E. 1975. *Mental health and going to school: The Woodlawn Program of assessment, early intervention, and evaluation.* Chicago: University of Chicago Press.

Koth, C. W., Bradshaw, C. P., and Leaf, P. J. 2009. Teacher Observation of Classroom Adaptation–Checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development* 42: 15 – 30.

Leaf, P. J., Schultz, D., Keys, S., and Ialongo, N. 2002. *The Teacher Observation of Classroom Adaptation–Checklist TOCA-C.* Baltimore: Johns Hopkins Center for the Prevention of Youth Violence.

Malcarne, V. 1983. Evaluating humane education: The Boston study. *Humane Education* 7: 12 – 13.

Nicoll, K., Trifone, C., and Samuels, W. E. 2008. An in-class, humane education program can improve young students' attitudes toward animals. *Society & Animals* 16: 45 – 60.

Petras, H., Chilcoat, H. D., Leaf, P. J., Ialongo, N. S., and Kellam, S. G. 2004. Utility of TOCA-R scores during the elementary school years in identifying later violence among adolescent males. *Journal of the American Academy of Child and Adolescent Psychiatry* 43: 88 – 96.

Pinheiro, J., Bates, D., Deb Roy, S. Sarkar, D., and the R Development Core Team 2013. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1 – 113.

Preston, H. C. 1928. Milestones in humane education. *Journal of Education* 107: 181.

R Core Team 2013.*R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900 051-07-0, Retrieved from http://www.R-project.org/.

Revelle, W., 2014. *psych: Procedures for Personality and Psychological Research*. Evanston, Illinois: Northwestern University. Retrieved from http://cran.r-project.org/web/packages/psych/.

Schaeffer, C. M., Petras, H., Ialongo, N., Masyn, K. E., Hubbard, S., Poduska, J., and Kellam, S. 2006. A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: Prediction to young adult antisocial outcomes. *Journal of Consulting and Clinical Psychology* 74: 500 – 510.

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., and Kellam, S. 2003. Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology* 6: 1020 – 1035.

Unnebrink K. and Windeler, J. 2001. Intention-to-treat: Methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Statistics in Medicine* 20: 3931 – 3946.

Unti, B. and DeRosa, B. 2003. Humane education: Past, present and future. In *The State of the Animals II: 2003*, 27-50, ed. D. J. Salem and A.N. Rowan. Washington, DC: Humane Society Press.

Vockell, E. 1979. *Measuring attitudes toward animal life: A technical report to American Humane*. Hammond, IN: Humane/Environmental Education Development, Inc.

Vockell, E. and Hodal, E. F. 1980. Developing humane attitudes: What does research tell us? *Humane Education* 4: 19 – 21.

Western Institute for Research and Evaluation and National Association for Humane and Environmental Education 1983. *Primary Attitude Scale*. Logan, UT: Author.

## Acknowledgements

The authors would like to thank the faculty, staff, and especially students of the participating schools for their kind involvement in this study. In addition, we thank Kimberly Korona, Christopher Parrucci, Meena Alagappan, and Bradley Goldberg of HEART's New York branch for the opportunity to conduct this investigation. Finally, we thank Robert Schwalb, also of HEART, for his outstanding work administering the instruments to the participants in Chicago.

Table 1

*Participating schools' overall student demographic information.*

| | | Ethnicity Percents | | | | | Percent Receiving Free or Reduced School Lunch | Percent Female |
|---|---|---|---|---|---|---|---|---|
| School | Location | African-American* | Asian or Pacific Islander* | European-American* | Hispanic | Other or Multiple Ethnicities* | | |
| A | Chicago | 1 | 0 | 44 | 53 | 2 | 59 | 52 |
| B | Chicago | 47 | 7 | 6 | 35 | 5 | 95 | 41 |
| C | New York City | 29 | 5 | 2 | 63 | 1 | 100 | 46 |
| D | New York City | 53 | 1 | 1 | 45 | 0 | 77 | 52 |

* Non-Hispanic

Table 2

*Marginal means, sample sizes, and standard errors of the means for students' teacher-reported TOCA-C prosocial behavior scores, TOCA-C disruptive behavior scores, and IAS total scores.*

| | | | | Pre-Program | | Post-Program | |
|---|---|---|---|---|---|---|---|
| TOCA-C Sub-Score | Term | Group | *n* | *M* | *SEM* | *M* | *SEM* |
| Prosocial Behavior | Treatment | Experimental | 119 | 21.65 | 0.38 | 25.14 | 0.37 |
| | | Control | 48 | 21.79 | 0.74 | 23.04 | 0.79 |
| | Gender | Male | 86 | 21.58 | 0.48 | 23.99 | 0.50 |
| | | Female | 81 | 22.80 | 0.49 | 25.12 | 0.49 |
| | School | A | 65 | 22.82 | 0.50 | 25.01 | 0.50 |
| | | B | 52 | 20.22 | 0.51 | 25.00 | 0.70 |
| | | C | 21 | 22.80 | 0.75 | 24.35 | 0.65 |
| | Total Sample | | 167 | 21.69 | 0.34 | 24.54 | 0.36 |
| Disruptiveness | Treatment | Experimental | 119 | 16.21 | 0.69 | 13.92 | 0.53 |
| | | Control | 48 | 15.38 | 1.08 | 13.81 | 0.74 |
| | Gender | Male | 86 | 16.93 | 0.92 | 14.35 | 0.67 |
| | | Female | 81 | 14.95 | 0.67 | 13.41 | 0.56 |
| | School | A | 67 | 14.12 | 0.60 | 13.28 | 0.54 |
| | | B | 54 | 19.22 | 1.40 | 15.93 | 0.99 |
| | | C | 46 | 14.85 | 0.77 | 12.39 | 0.65 |
| | Total Sample | | 167 | 15.97 | 0.58 | 13.89 | 0.44 |
| | Treatment | Experimental | 62 | 100.71 | 0.69 | 110.41 | 0.72 |
| | | Control | 173 | 99.42 | 1.07 | 103.64 | 1.01 |
| | Gender | Male | 119 | 98.63 | 0.80 | 106.62 | 0.88 |
| | | Female | 116 | 102.15 | 0.82 | 110.68 | 0.86 |
| IAS | School | A | 65 | 105.65 | 1.14 | 112.49 | 1.17 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | B | 49 | 100.51 | 1.28 | 108.31 | 1.46 |
| | C | 56 | 95.62 | 0.86 | 106.79 | 1.35 |
| | D | 65 | 99.08 | 0.98 | 16.58 | 0.96 |
| | Total Sample | 235 | 100.37 | 0.58 | 108.62 | 0.63 |

*n* = model group sample size; *M* = marginal mean; *SEM* = standard error of the marginal mean

Table 3

*Summary of terms in the multi-level change model predicting students' standardized TOCA-C* prosocial *scores.* N = *167.*

| | | Model | | |
|---|---|---|---|---|
| | | Time-Only | Main Treatment Added | Time x Treatment Added |
| Time | β | 0.073 | 0.200 | 0.023 |
| | *p* | 0.032 | < 0.001 | 0.733 |
| Treatment | β | | 0.304 | 0.980 |
| | *p* | | 0.005 | < 0.001 |
| Treatment x Time | β | | | 0.665 |
| | *p* | | | < 0.001 |
| $-2LL$ | | 1134.04 | 1127.54 | 1093.86 |
| Difference in $-2LL$* | | | 6.50 | 33.68 |

* Critical value: $\chi^2_1$ = 3.84 (at α = 0.05) for tests between adjacent models; $\chi^2_2$ = 5.99 for model 3 versus model 1

Table 4

*Summary of terms in the multi-level change model predicting students' standardized TOCA-C* disruptiveness *scores.* N = *167.*

| | | Model | | |
|---|---|---|---|---|
| | | Time-Only | Main Treatment Added | Time x Treatment Added |
| Time | β | −0.055 | −0.174 | −0.113 |
| | *p* | 0.077 | < 0.001 | 0.072 |
| Treatment | β | | 0.320 | 0.521 |
| | *p* | | 0.004 | 0.002 |
| Treatment x Time | β | | | −0.177 |
| | *p* | | | 0.100 |
| $-2LL$ | | 1103.97 | 1095.79 | 1093.05 |
| Difference in $-2LL$* | | | 8.19 | 2.73 |

* Critical value: $\chi^2_1$ = 3.84 (at α = 0.05) for tests between adjacent models; $\chi^2_2$ = 5.99 for model 3 versus model 1

Table 5

*Summary of terms in the multi-level change model predicting students' standardized IAS scores.* N = *236.*

| | | Model | | |
|---|---|---|---|---|
| | | Time-Only | Time + Treatment | Time + Treatment + Time*Treatment |
| Time | β | 0.172 | 0.379 | 0.033 |
| | *p* | < 0.001 | < 0.001 | 0.548 |
| Treatment | β | | 0.496 | 1.325 |
| | *p* | | < 0.001 | < 0.001 |
| Treatment x Time | β | | | 0.976 |
| | *p* | | | < 0.001 |
| $-2LL$ | | 1435.10 | 1418.73 | 1303.14 |
| Difference in $-2LL$* | | | 16.38 | 115.59 |

* Critical value: $\chi^2_1$ = 3.84 (at α = 0.05) for tests between adjacent models; $\chi^2_2$ = 5.99 for model 3 versus model 1

Figure 1: TOCA-C prosocial score marginal means and 95% confidence intervals for the experimental and control groups at pre- and post-program.

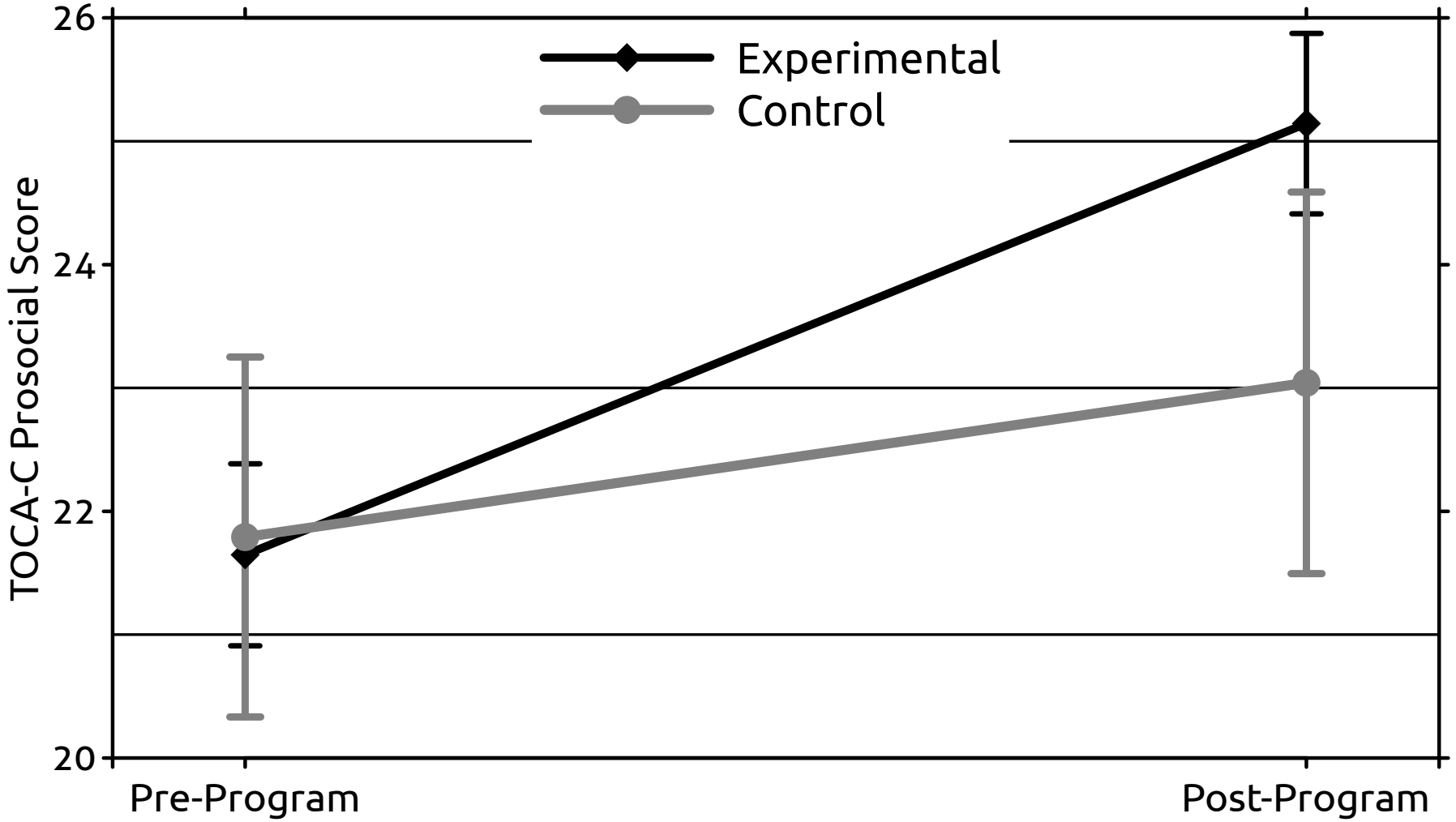

Figure 2 TOCA-C disruptiveness score marginal means and 95% confidence intervals for the experimental and control groups at pre- and post-program.

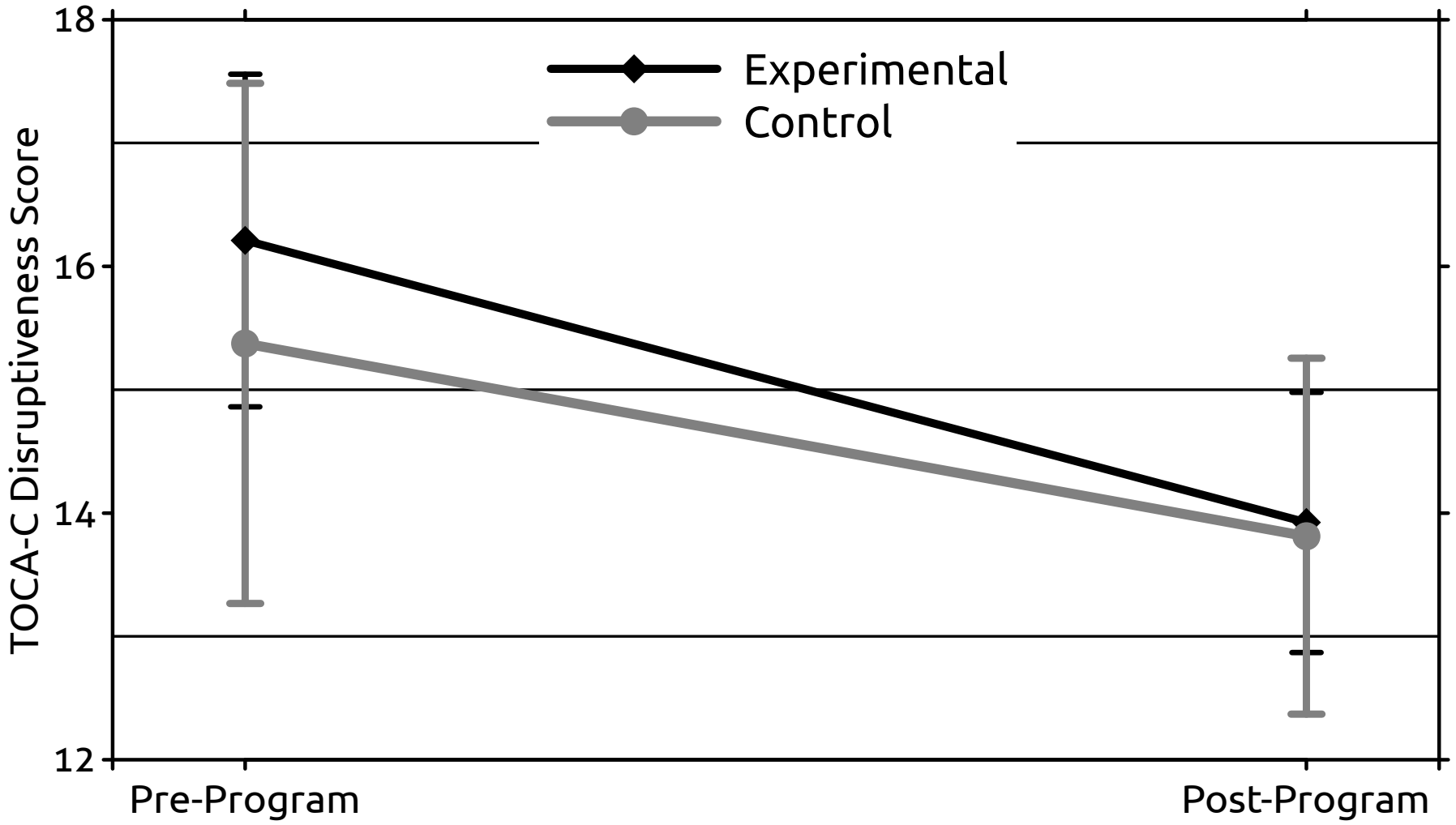

Figure 3: IAS score marginal means and 95% confidence intervals for the experimental and control groups at pre- and post-program.