

Who are You Going to Believe, Me or Your Own Eyes: The Use of Pre-Post Test Designs to Evaluate Effectiveness of Autism Treatments

By Thomas Zane, Ph.D., BCBA-D

Center for Applied Behavior Analysis at The Sage Colleges

A prerequisite to accessing clinical treatment for persons with Autism Spectrum Disorders is to first understand what treatments might be effective and have a chance of delivering positive results. And a prerequisite to determining what treatments have actually been proven to be “evidenced-based” is understanding what allows a research study to give valid and believable results.

Consider a study recently published by Rossignol and Rossignol (2006), in which they assessed the effect of a hyperbaric oxygen chamber on a range of symptoms of six children diagnosed with autism. Prior to starting the hyperbaric oxygen therapy (HBOT), the researchers assessed the participants on three measures, the Autism Treatment Evaluation Checklist, the Childhood Autism Rating Scale, and the Social Responsiveness Scale. The children then participated in HBOT for 40, 1-hour sessions, and the researchers then re-assessed the participants using the same measures as in the pretest. For most of the children, the post-test scores were lower on each assessment (for these instruments, a lower score suggests fewer symptoms of autism and improved functioning). The authors suggested that the HBOT was responsible for the improvement.

Consider a study by Gutstein, Burgess, and Monfort (2007), in which they assessed the effectiveness of Relationship Development Intervention, an autism treatment. The authors selected 16 children with autism and reviewed their files, noting their test scores on various measures prior to receiving RDI - subsets of the Autism Diagnostic Observation

Schedule and Autism Diagnostic Interview-Revised. Additionally, parents provided information about each child's educational placement (on a continuum of intrusiveness) and level of "flexibility" (i.e., child's comfort level reacting to change in his/her life and routine). After obtaining these measures, the participants received RDI for an average of 18 months. Following treatment, Gutstein, et al. conducted post-test assessments using the same measures as the pretest. For most children, the authors concluded that the post-test scores improved over pre-test scores, and suggested that RDI was responsible for the improvement.

Researchers and clinicians often attempt to demonstrate the effectiveness of an autism treatment by using this common "pre-post" test design (also called "before-after," "AB," and "one-group, pretest-posttest design; e.g., Drew, Hardman, & Hosp, 2008; Fraenkel & Wallen, 2009). The general strategy in a pre-post test study is to recruit one group of subjects, obtain some measurement of the critical dependent variable(s) hypothesized to be changed by the treatment, implement the treatment protocol, and then re-administer the same measurement as pretest. There is an assumption that if the post-test scores have changed positively from the pretest scores, then the change is due to the treatment. Many researchers and treatment developers use this basic design (e.g., Krantz, 2009; Linderman & Steward, 1999; Rossignol, Rossignol, James, Melnyk, & Mumper, 2007).

The important question is, does this design provide convincing proof that the treatment caused the improvement in the variable(s) being measured? The answer is unambiguous - this basic design never permits confirmation of cause and effect between

the treatment and positive changes in the dependent measures(e.g., autism symptomology; Drew, Hardman, & Hosp, 2008; Gay, Mills, & Airasian, 2009).

The weakness of this design (to demonstrate causal relationships) relates to its inability to minimize “internal validity” threats. The internal validity of a research study refers to the level of confidence in believing that changes in the variables being measured are due to the treatment protocol being used. If the research study is designed to eliminate any explanation other than the treatment changing what is being measured, then that study has strong internal validity. On the other hand, if the research study is designed in a way that allows explanations other than the treatment variable to possibly be influencing what is being measured, then that study will have weak internal validity, and the conclusion must be that the treatment may not be the only reason for the change in the dependent measurements. And if there is an assumption that variables other than the treatment could have produced the changes in what is measured, one must conclude that the treatment probably did not cause the changes.

The pretest-posttest design is fatally flawed with respect to internal validity. For example, if participants improve from pretest to posttest, the improvement could be due to simply the participants maturing (physically or psychologically) over the course of the experiment. Consider a research project done over the course of a year with preschoolers with autism. An improvement in assessment from pretest to posttest (after one year) could be due simply to the natural maturation of the participants, rather than influence of the treatment. Another possible threat to believing that a treatment caused any positive changes relates to participants who were chosen on the basis of extremely low scores (or

extremely low performance) on the variable(s) being measured in the pretest. Generally, extremely low scores will often improve, and extremely high scores will often decline, given repeated assessments, just because they are so extreme. Thus, any study that involves participants because they scored very low or very high on the dependent measures, and that uses a pretest-posttest design, is open to this particular threat and thus one cannot believe that the treatment caused any improvement.

The one group pretest-posttest design is flawed by several additional internal validity threats not discussed here. The reality is that any attempt to demonstrate the effectiveness of a treatment by using a pretest on one group of participants, then applying a treatment, followed by a reassessing the variables being tracked, will always be open to skepticism of linking improvement to treatment. This type of design will never allow strong confidence in the belief of a cause and effect connection between treatment and improvement.

All research is not equal in quality. Just because a research study has been conducted and shows positive changes in some aspects of autism does not necessarily mean that the treatment was responsible for those changes. Since autism is said by some to be a “fad magnet” (e.g., Jacobson, Foxx, & Mulick, 2005), parents and other consumers must critique any research study that purports to show a positive effect of a treatment, and try to determine if the positive changes could be due to other explanations, or could only be due to the treatment. By activating their “baloney detectors” (Sagan, 1999), parents, care givers, and service providers can avoid adopting treatments that have no proof of effectiveness, and thus be more likely to embrace treatments for which there is a body of well-designed

research supporting a cause and effect relationship. Research in autism treatments that purportedly shows evidence of effectiveness, but that utilizes only pretest-posttest studies, needs to be viewed with caution and must not be thought of as producing valid conclusions that allow consumers and caregivers to believe that the treatment in fact works. Accessing clinical treatment services could be enhanced by better understanding of the flaws in this basic and commonly used research design.

References

- Drew, C. J., Hardman, M. L., & Hosp, J. L. (2008). *Designing and conducting research in education*. Thousand Oaks, California: Sage Publications, Inc.
- Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education. Seventh edition*. New York: McGraw-Hill.
- Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational research: Competencies for analysis and applications, Ninth Edition*. Upper Saddle River, NJ: Pearson.
- Gutstein, S. E., Burgess, A. F., & Montfort, K. (2007). Evaluation of the Relationship Development Intervention program. *Autism*, 11, 397-412.
- Jacobson, J. W., Foxx, R. M., & Mulick, J.A. (2005). *Controversial therapies for developmental disabilities: Fad, fashion, and science in professional practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Krantz, S. (2009). Craniosacral therapy: Helping improve brain function. Retrieved March 8, 2010 at <http://www.upledger.com/content.asp?id=76&mid=2>.
- Linderman, T. M., & Steward, K. B. (1999). Sensory integrative-based occupational therapy and functional outcomes in young children with pervasive developmental disorders: A single subject study. *American Journal of Occupational Therapy*, 53, 207-213.
- Rossignol, D. A., & Rossignol, L. W. (2006). Hyperbaric oxygen therapy may improve symptoms in autistic children. *Medical Hypotheses*, 67, 216-228.

Rossignol, D. A., Rossignol, L. W., James, S. J., Melnyk, S., & Mumper, E. (2007). The effects of hyperbaric oxygen therapy on oxidative stress, inflammation, and symptoms in children with autism: An open-label pilot study. *MBC Pediatrics*. Retrieved on March 8, 2010 at <http://www.biomedcentral.com/1471-2431/7/36>.

Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark*. New York, NY: Random House.