

The College at Brockport: State University of New York Digital Commons @Brockport

Education and Human Development Faculty
Publications

Education and Human Development

12-2010

An Analysis of Evidence-Based Practices in the Education of Learners with Autism Spectrum Disorders

Michael R. Mayton

West Virginia University, michael.mayton@mail.wvu.edu

John J. Wheeler

Western Michigan University


Anthony L. Menendez

Cleveland State University

Jie Zhang

The College at Brockport, jzhang@brockport.edu

Follow this and additional works at: https://digitalcommons.brockport.edu/ehd_facpub

 Part of the [Disability and Equity in Education Commons](#), and the [Special Education and Teaching Commons](#)

Repository Citation

Mayton, Michael R.; Wheeler, John J.; Menendez, Anthony L.; and Zhang, Jie, "An Analysis of Evidence-Based Practices in the Education of Learners with Autism Spectrum Disorders" (2010). *Education and Human Development Faculty Publications*. 14. https://digitalcommons.brockport.edu/ehd_facpub/14

Citation/Publisher Attribution:

Mayton, M. R., Wheeler, J. J., Menendez, A. L., & Zhang, J. (2010). An analysis of evidence-based practices in the education of learners with autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities*, 45(4), 539-551.

This Article is brought to you for free and open access by the Education and Human Development at Digital Commons @Brockport. It has been accepted for inclusion in Education and Human Development Faculty Publications by an authorized administrator of Digital Commons @Brockport. For more information, please contact kmyers@brockport.edu.

An Analysis of Evidence-Based Practices in the Education and Treatment of Learners with Autism Spectrum Disorders

Michael R. Mayton
West Virginia University

John J. Wheeler
Western Michigan University

Anthony L. Menendez
Cleveland State University

Jie Zhang
The College at Brockport: State University of
New York

Abstract: Horner et al. (2005) present a review substantiating how single-subject research methodology can be utilized to determine whether interventions are evidence-based practices (EBPs). The current study utilized the Horner et al. research piece to: (a) systematically identify a set of quality standards for the evaluation of single-case research methodology used with learners with autism spectrum disorders (ASD), (b) operationalize these standards for evaluators, (c) investigate three additional quality indicators related to external validity (multiple studies, locations, and researchers), (d) create a protocol for evaluators, and (e) gather and analyze data from studies that meet a set of predefined criteria. Published in seven journals across ten years, identified studies (N = 160) were subjected to an analysis across 23 EBP quality standards that revealed increasing compliance with EBP standards over time yet persistent and pervasive difficulty in adequately satisfying at least five indicators integrally tied to external validity.

The term EBP originated in the field of medicine as early as the 1970s (Odom et al., 2005) and has become an integral part of medical education. The intent of this practice was to minimize the gap between research and practice with the end goal being directed toward the use of scientific evidence as the method of choice for physicians in the diagnosis and treatment of illness. The movement for evidence-based practices (EBP) in the field of education was introduced with the passage of the No Child Left Behind (NCLB) Act of 2001 (Odom et al.). This significant piece of legislation emphasized the use of science-based methods by teachers in their classrooms and was aimed at restructuring the educational system with increased standards and accountability as its cornerstones (Simpson, 2005). The role of EBP in the field of special education has certainly been influenced in part by the NCLB Act, yet the field has had a rich

history of research-based methods in the education and treatment of persons with disabilities (Odom et al.). Many of the early investigations in the field were derived from case studies that were generated from the traditions in the field of psychology. As the field of special education progressed, we witnessed the use of more elaborate research designs, including the use of single-subject experimental designs that resulted in more experimentally valid research. Yet, despite this rich learning history as a field, special education, like the field of education in general, has been challenged with a significant gap between research and practice. In other words, what we know to be effective in terms of teaching and learning based on research findings is often not employed with learners in the classroom. This gap is also evident with respect to the education and treatment of learners with autism, which has resulted in the term EBP gaining a great deal of recent notoriety with respect to the treatment of autism, especially given the increasing prevalence rates of children being diagnosed with the condition. Now, 1 in every 110 children is estimated to be diagnosed with autism (Centers for Disease

Correspondence concerning this article should be addressed to Michael R. Mayton, West Virginia University, Department of Special Education, 508-C Allen Hall, Morgantown, WV 26506-6122.

Control, 2009). Children and families affected by autism represent a very vulnerable population, and understandably families seek out a variety of treatments for their children in an effort to minimize the lifelong effects of autism. Conversely, the availability of non-evidence-based treatments for children affected by autism are substantial in number and often pose a threat to the well-being of these children and their families (Simpson), thus the importance of identifying evidence-base treatments within research to better inform practice. To better understand EBP, one must understand at least one prominent way in which the term has already been operationalized for learners with autism spectrum disorders (ASD).

The National Professional Development Center (NPDC) on Autism Spectrum Disorders (2009) defines EBP in the context of peer-reviewed research published in professional journals. The NPDC's position is that, in order to be termed an EBP, an intervention must be presented within studies published in such journals and constitute a research base substantiated by: (a) two high quality experimental or quasi-experimental group design studies, (b) five high quality single-subject design studies conducted by three different researchers, or (c) one high quality randomized or quasi-experimental group design study and three high quality single-subject design studies conducted by at least three different researchers.

The importance of selecting and using EBP for learners with ASD cannot be overstated, especially given (a) the prevalence of ASD is increasing at an alarming rate, (b) the high-stakes nature of intervention outcomes for children with ASD and their families, and (c) the history of questionable treatments that have been marketed to children and families affected by autism. These are obvious justifications for the importance of EBP in the research literature and subsequent practice. However, a gap not only exists between researchers and practitioners but in how research is conducted in terms of quality assurance. Issues such as treatment fidelity (i.e., if interventions are implemented in a reliable manner, as designed) have plagued the literature. One study conducted by Wheeler, Baggett, Fox, and Blevins (2006) reviewed behav-

ioral treatment studies in the field of autism from 1993–2003 across nine peer reviewed journals specializing in such studies. The results from this investigation revealed that of the 60 studies reviewed, only 11 studies assessed treatment integrity and provided an adequate operational definition of the independent variable. This fact is startling given how research can potentially impact practice.

The purpose of this study was to provide an analysis of research relative to EBP with learners identified with ASD, using the standards for evaluating single-subject research as recommended by Horner et al. (2005). Implications for practice and research are provided.

Method

The Evidence-Based Practice Construct

Horner et al. (2005) present a review and rationale substantiating the foundation that existing scientific knowledge and procedure regarding single-subject research methodology can be utilized to determine whether interventions are EBPs. The authors discuss, define, and provide examples of eight quality indicators within single-subject research and add further recommendations for strengthening the external validity of this research. The current study utilized the Horner, et al. research piece to: (a) systematically identify a set of quality standards for the evaluation of single-case research methodology used with learners with ASD, (b) operationalize these standards for evaluators, (c) investigate three additional quality indicators related to external validity, (d) create a protocol for evaluators, and (e) gather and analyze data from studies that meet a set of predefined criteria.

Journals and Article Selection Criteria

Seven journals were chosen for the review based on their prominence in the field of single-subject research conducted with persons with ASD. Selected journals were as follows: (a) *Education and Training in Developmental Disabilities* (ETDD), (b) *Education and Treatment of Children* (ETC), (c) *Focus on Autism and Other Developmental Disabilities* (FOCUS), (d) *Journal of Applied Behavior Analysis*

(JABA), (e) *Journal of Autism and Developmental Disorders* (JADD), (f) *Journal of Positive Behavior Interventions* (JPBI), and (g) *Research in Developmental Disabilities* (RDD).

In order to be included in the current study, research had to adhere to four main criteria. First, all participants included in a study had to have a diagnosis that is somewhere on the autism spectrum of disorders. Articles utilizing "mixed" samples were rejected. A mixed sample was considered to be one in which one or more of the participants in a study had a sole diagnosis other than autism. For example, studies were rejected if they utilized samples that included participants with a sole diagnosis of mental retardation as well as participants with autism. However, studies were accepted in which participants were diagnosed with autism and comorbid disorders (e.g., one or more individuals were diagnosed with *both* autism and mental retardation). In addition, all participants had to be from 3 to 21 years of age, and each study had to involve a behavioral treatment. A behavioral treatment was defined as an intervention-based study aimed at the development of skills or behaviors within participants using a single-case experimental design. Last, all articles had to have been published within the predefined, ten-year range (2000–2009).

Standards and Operational Definitions

Twenty-three standards were distilled directly from the Horner et al. (2005) discussion of quality indicators found within single-subject research. Horner, et al. also recommended that external validity be strengthened "through systematic replication of effects across multiple studies conducted in multiple locations and across multiple researchers" (p.171), hence the current examination of: (a) most prolific authors, (b) locations where research is being conducted, (c) number of participants per study, and (d) the number of studies within each category of independent variable. Table 1 provides a complete list of the indicators and standards, as well as examples of operational definitions.

Raters, Ratings, and Interobserver Reliability

The authors of the current study served as raters due to their expertise and applied experience in the area of single-subject research design. The first author provided the other authors with: (a) journal assignments, (b) the Horner et al. (2005) article, (c) article selection criteria, and (d) the data collection protocol (including general instructions, item descriptions, operational definitions, and indexed references to specific content within the Horner, et al. article). Raters were given a two-week period for review of materials and clarification of any questions, and during this period the first author engaged raters in a series of phone and face-to-face conversations to check for understanding of the protocol, associated materials, and procedures.

Raters were instructed to evaluate the 23 standards across each study in terms of "acceptable" or "not acceptable." In order for a study to be rated "acceptable" within a particular standard, the study had to present evidence fully satisfying each part of the operational definition for that standard. No partial ratings were allowed. For instance, in order to rate a study as one that describes the independent variable (IV) with replicable precision, raters had to find adequate evidence that matched the operational definitions for: (a) systematic manipulation of the IV, (b) IV under the control of the experimenter, and (c) measurement description and reporting of treatment integrity data. Failure in one or more of these areas would result in a rating of "not acceptable" for this indicator.

Once all initial ratings were completed, 20% of the articles chosen for review within the current study ($n = 32$) were randomly selected and distributed to the first and second authors, making sure that the authors received only articles that had been previously rated by other reviewers. The studies were then independently read and rated, and agreement was calculated by dividing the number of agreements plus disagreements and multiplying by 100. The percentage of agreement across all ratings averaged 93.5% (23 ratings for each of 32 studies = 736 individual ratings).

TABLE 1

Quality indicators in single-subject research, rating standards, and example operational definitions.

<i>Quality Indicator (Horner, et al., 2005)</i>	<i>Rating Standard</i>	<i>Example Operational Definition</i>
Description of participants and settings	<ol style="list-style-type: none"> 1. Description of participants 2. Participant selection process 3. Description of setting 	<p><u>Description of participants (p. 167):</u> <i>Participants are described in sufficient detail to allow others to select individuals with similar characteristics.</i></p> <ol style="list-style-type: none"> 1. The specific disability and the specific instrument and process used to determine the disability are identified. 2. Specific characteristics directly pertaining to the DV and IV are reported (e.g., if the DV and IV deal with some aspect of communication, prior levels of communication ability should be specified).
Dependent variable	<ol style="list-style-type: none"> 4. Operational definition 5. Measurement procedure 6. Valid and precise 7. Repeated measurement 8. Interobserver agreement 9. Social significance 	<p><u>Interobserver agreement (p. 167):</u> <i>IOA is reported to a sufficient extent and meets minimal or higher standards.</i></p> <ol style="list-style-type: none"> 1. IOA is assessed across each variable, participant, and condition within the study. 2. These data meet or exceed minimal standards: IOA = 80% and Kappa = 60%.
Independent variable	<ol style="list-style-type: none"> 10. Description 11. Systematic manipulation 12. Implementation fidelity 	<p><u>Implementation fidelity (a.k.a., treatment integrity, treatment fidelity) (p. 168):</u> <i>The fidelity of IV implementation is fully documented.</i></p> <ol style="list-style-type: none"> 1. Overt measurement of implementation fidelity reportedly occurred. 2. These data were reported.
Baseline	<ol style="list-style-type: none"> 13. Pattern of responding 14. Description of conditions 	<p><u>Pattern of responding (p. 168):</u> <i>Measurement of the DV during the baseline condition demonstrates a pattern of responding that allows prediction of future responding.</i></p> <ol style="list-style-type: none"> 1. There are at least five data points displayed in the initial baseline condition (see below for exceptions). 2. The data display no discernable trend, or they display a trend in the opposite direction than that predicted by the intervention.

Results

General Study Characteristics

Articles from JABA, JPBI, JADD, and FOCUS made up the majority (79%) of the 160 studies included in the current study. Across all included articles, no study exceeded 10 participants, with the next highest maximum being 7 participants. Three journals published studies that were conducted

within all six of the major United States geographical areas (as well as outside the U.S.), and across all journals multicomponent independent variables (IVs) were the most commonly found. IVs that were composed of treatment packages, or those made up of more than one significant component, were labeled “multicomponent.” This was done to avoid the potential confound of including studies within a particular evi-

TABLE 1 (Continued)

Quality Indicator (Horner, et al., 2005)	Rating Standard	Example Operational Definition
Experimental control/Internal validity	15. Demonstrated effect 16. Controls for threats 17. Demonstration of experimental control	<p><u>Demonstration of experimental control (p. 168):</u> <i>Experimental control is demonstrated to the extent that a functional relationship is indicated between the manipulation of the IV and the change in the DV.</i></p> <ol style="list-style-type: none"> 1. The design documents three demonstrations of experimental effect at three different points in time with a single participant, or across different participants. 2. Documentation of experimental control is achieved through: (a) the introduction and withdrawal of the IV; (b) the staggered introduction of the IV at different points in time; or (c) the iterative manipulation of the IV (or levels thereof) across observation periods.
External validity	18. Replication	<p><u>Replication (p. 171):</u> <i>Experimental effects are replicated across relevant variables in order to increase external validity.</i></p> <p>Experimental effects are demonstrated across at least three different participants, settings, sets of materials, and/or behaviors.</p>
Social validity	19. Social importance of the DV 20. Magnitude of change 21. Practical/cost-effective 22. Enhancement	<p><u>Enhancement (p. 172):</u> <i>Social validity is enhanced by social importance, typical application, and the positive report of intervention agents.</i></p> <ol style="list-style-type: none"> 1. DV(s) were selected that have a high “social importance,” or emphasis on helping an individual effectively operate in typical/needed contexts in which other people also operate. 2. It was demonstrated that the intervention(s) could be implemented with fidelity by typical intervention agents (e.g., parents) in typical contexts across meaningful periods of time. 3. Typical intervention agents report intervention procedures to be feasible within available resources, effective, and useful after formal intervention is discontinued.
Research question(s)	23. Appropriateness of research question(s) for single-subject research methods	<p><u>Appropriateness of research question(s) (p. 172):</u> <i>Single-subject methodology is used to investigate functional relations between the IV(s) and DV(s), pertaining to the performance of a specific individual.</i></p> <ol style="list-style-type: none"> 1. Each question deals with the effects that systematic manipulation of an IV (or component of an IV) has on one or more DV(s). 2. Each question is concerned with/focuses upon the performance of specific individuals under a given set of conditions.

dence-base grouping (e.g., “social stories”) that have other significant intervention features that could be responsible for some or all of the reported results. See Table 2 for details regarding these characteristics.

Indicators of External Validity

Replication of effects. Most of the articles identified (80%) had from one to three participants, as expected within single-subject re-

TABLE 2

General study characteristics.

<i>Journal</i>	<i>Number and Percent of Studies Identified (N = 160)</i>	<i>Maximum Participants Across Studies</i>	<i>Geographical Locations Represented</i>	<i>Geographical Locations Unrepresented</i>	<i>Most Common Independent Variable(s)</i>
ETDD	12 (8%)	6	Midwest Mountain states Non-U.S. South	Atlantic states Pacific states Southwest	multicomponent
ETC	6 (4%)	4	Atlantic states Pacific states South	Midwest Mountain states Non-U.S. Southwest	multicomponent
FOCUS	22 (14%)	6	Atlantic states Midwest Mountain states Non-U.S. Pacific states South Southwest		social stories, multicomponent
JABA	50 (31%)	7	Atlantic states Midwest Mountain states Non-U.S. Pacific states South Southwest		multicomponent, prompting/reinforcement
JADD	26 (16%)	10	Atlantic states Midwest Mountain states Non-U.S. Pacific states South Southwest		multicomponent, social stories
JPBI	29 (18%)	4	Atlantic states Midwest Non-U.S. Pacific states South	Mountain states Southwest	video modeling, multicomponent
RDD	15 (9%)	6	Atlantic states Midwest Non-U.S. Pacific states South Southwest	Mountain states	prompting/reinforcement, imitation training

search studies. However, it was demonstrated that a sole single-subject study could effectively utilize as many as 10 participants, though this occurred in only one study that fit

the search criteria (see Figure 1 for percentages within each category).

In addition to reviewing the number of participants within studies, replication of effects

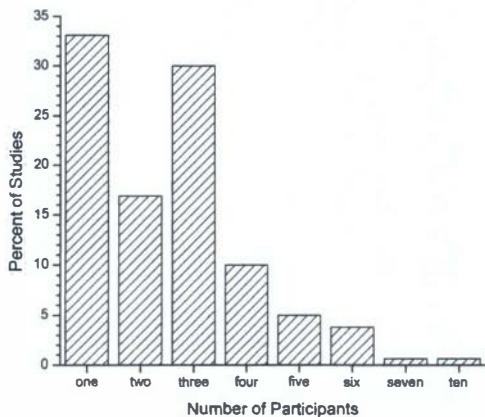


Figure 1. Percent of studies by number of participants.

was examined by categorizing the types of IVs utilized across studies. The formation of categories was based on the following decision guidelines: (a) there had to be at least two studies with the same IV, and (b) the IV could not be classified as “multicomponent” (as previously described). The top four IV categories in terms of frequency across the ten-year examination period were: (a) multicomponent,

(b) prompting/reinforcement, (c) social stories, and (d) video modeling, comprising 64% of the articles across all categories. Table 3 provides a complete listing of the categories that emerged using the decision guidelines.

Multiple locations. Within groupings of the six major U.S. geographical areas and non-U.S. locations, the largest numbers of studies that fit the search criteria were produced in the South and Atlantic States regions (together, 44% of all studies). The regions producing the fewest studies were the Southwest and Mountain States (together, 12% of all studies). Figure 2 graphically displays percentages across these regions.

Multiple researchers. Using the criterion of identifying first-authors who published 3 or more studies within the ten-year examination period, only six authors were categorized as “most prolific.” Approximately 14% of the articles identified for use within the current study ($n = 22$) were produced by these six authors, with one author producing 5 studies (3.1% of all studies identified), two authors producing 4 studies each (2.5% each), and three authors producing 3 studies each (1.8% each).

TABLE 3
Number of studies by independent variable (IV).

<i>Independent Variable (brief descriptor)</i>	<i>Number of Studies: 2000–2009</i>
1. Activity schedule.....	4
2. Computer-based instruction (CBI)	3
3. Choice	3
4. Differential observing responses (DOR)	2
5. Discrete trial training (DTT)	3
6. Functional communication training (FCT)	7
7. Imitation training.....	5
8. Language instruction.....	3
9. Modeling.....	2
10. Multicomponent	35
11. Picture Exchange Communication System (PECS).....	4
12. Peer-mediated/assisted interventions.....	5
13. Prompting/reinforcement	23
14. Scripts.....	3
15. Social stories	14
16. Strategy instruction	3
17. Video modeling.....	13

Note. IVs that appeared in only one study across the ten-year period were not included in this analysis.

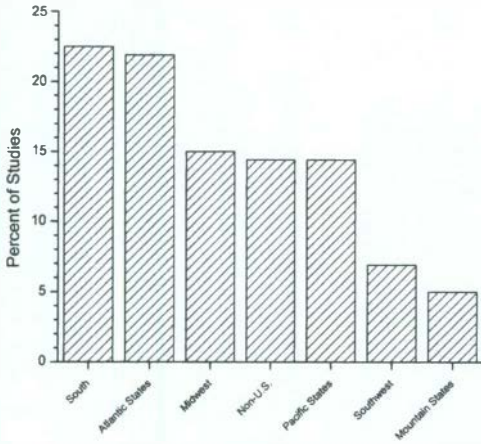


Figure 2. Percent of studies by geographical location.

Standards Ratings

Standards ratings by IV. The four IV categories with the highest percentages of “not acceptable” ratings were: (a) discrete trial training (41%), (b) choice (28%), (c) video modeling (22%), and (d) functional communication training (20%). The four IV categories with the highest percentages of “acceptable” ratings were: (a) strategy instruction (96%), (b) activity schedules (95%), (c) computer-based instruction (93%), and (d) imitation training (92%). Figure 3 provides a graphical representation of these percentages across all IV categories identified, and Table 4 shows percentage groupings across individual IVs.

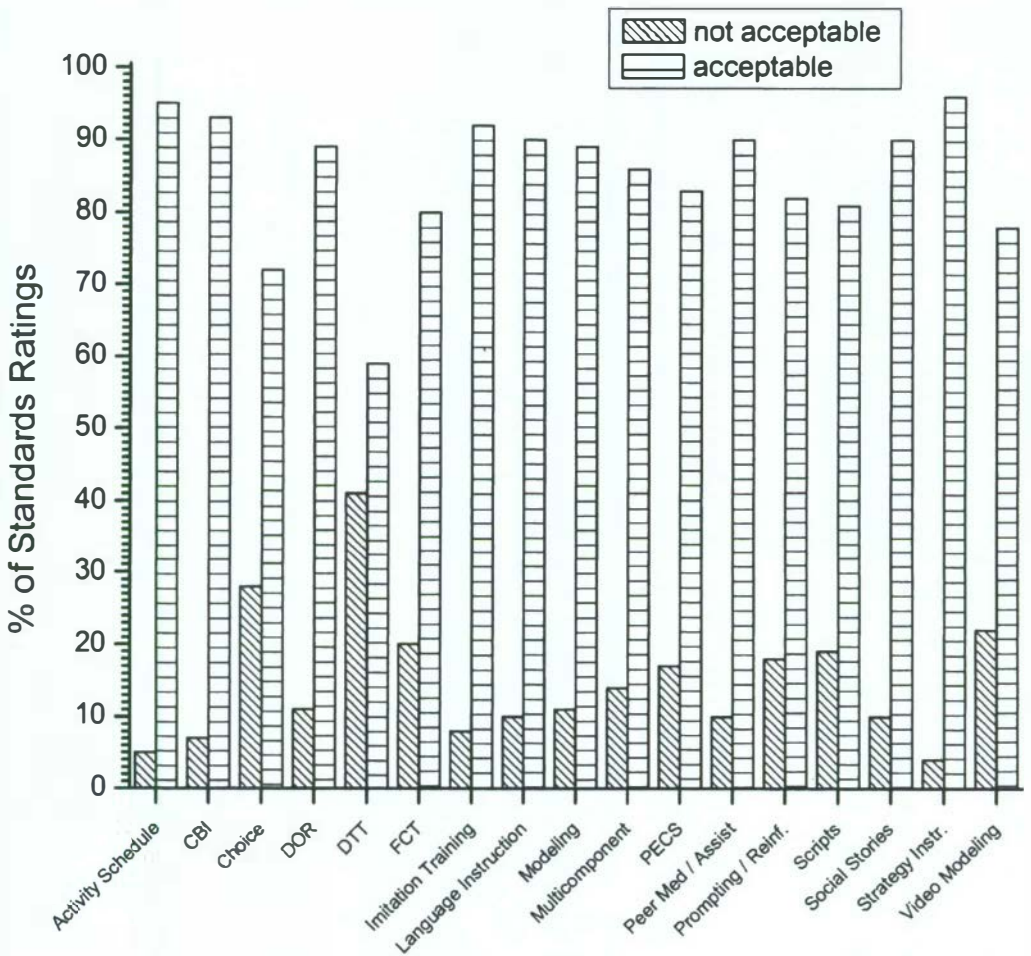


Figure 3. Percentage of standards ratings by IV.

TABLE 4

Grouped standards ratings across IVs.

Standard	Activity sched.	CBI	Choice	DOR	D'IT	FACT	Imitation train.	Language instr.	Modeling	Multicomponent	PECS	Peer-med./ asst.	Prompting/R.	Scripts	Social Stories	Strategy Instr.	Video Modeling
1	/	/	⊗	⊗	⊗	○	/	/	⊗	○	/	/	●	⊗	/	/	/
2	/	○	/	⊗	⊗	●	/	/	⊗	●	/	/	/	/	/	⊗	/
3	/	/	⊗	⊗	⊗	○	/	/	●	/	/	/	/	/	/	/	/
4	/	/	○	/	○	○	/	/	/	/	/	/	/	/	/	/	/
5	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
6	/	/	○	/	○	/	/	/	/	/	/	/	/	/	/	/	/
7	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
8	/	/	○	/	○	/	/	/	/	/	/	/	/	/	/	/	/
9	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
10	/	/	○	/	○	/	/	/	/	/	/	/	/	/	/	/	/
11	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
12	⊗	○	⊗	●	⊗	●	/	/	⊗	⊗	●	○	●	⊗	/	/	⊗
13	/	/	○	/	○	/	/	/	/	/	/	/	/	/	/	/	/
14	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
15	/	/	○	/	○	/	/	/	/	/	/	/	/	/	/	/	/
16	/	/	●	/	⊗	/	/	/	/	/	/	/	/	○	/	/	/
17	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
18	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/
19	/	○	/	/	⊗	/	/	/	/	/	/	/	/	/	/	○	/
20	/	/	⊗	/	○	/	/	/	/	/	/	/	/	○	/	/	○
21	/	○	○	/	○	/	/	/	/	/	/	/	/	○	/	/	/
22	/	○	○	/	○	/	/	/	/	/	/	/	/	⊗	/	/	/
23	/	/	/	/	○	/	/	/	/	/	/	/	/	/	/	/	/

Note :

- / = 80% (or higher) ratings of "acceptable" across studies
- = 70% (down to 51%)
- ⊗ = 50% (down to 31%)
- = 30% (or below)

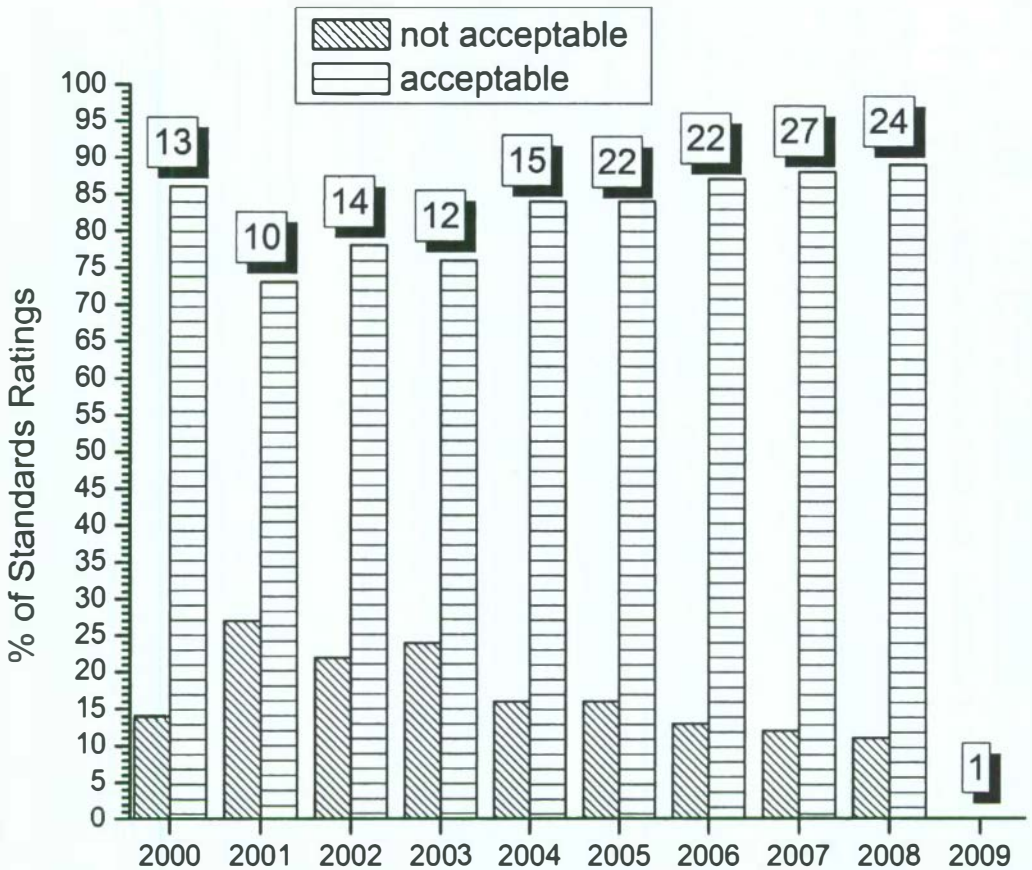


Figure 4. Percentage of standards ratings by year. Above each set of bars is the number of studies identified within the corresponding year. Note that at the time of the current study, only one article was identified from 2009, and therefore no comparable examination of standards ratings could take place for that year.

Standards ratings by year. Ratings partitioned out by years reveal that: (a) the number of studies conducted per year has steadily increased, and (b) the percentage of “acceptable” standards ratings within the current study also steadily increased. Figure 4 displays these percentages across the ten-year examination period.

Standards ratings by journal. Ratings of “acceptable” across all journals fell below 80% in five areas: (a) description of participants (61%), (b) participation selection process (40%), (c) description of setting (71%), (d) implementation fidelity (42%), and (e) controls for threats (78%). Approximately 16% of all ratings across all journals were in the “not acceptable” category (see Figure 5 for percentages of both ratings across all standards and journals).

Ratings by individual journals varied widely across the 23 standards. However, a pattern of “not acceptable” ratings can still be seen in standards 1 (description of participants), 2 (participant selection process), and 12 (implementation fidelity; all as discussed above), and all journals received standards ratings at 80% or above in a number of areas (from 48% to 91% of the 23 standards). Table 5 displays “acceptable” standards ratings by individual journal and percentage groupings.

Summary and Conclusions

The purpose of this review was to provide an analysis of EBP within single-subject research conducted with learners with ASD by using

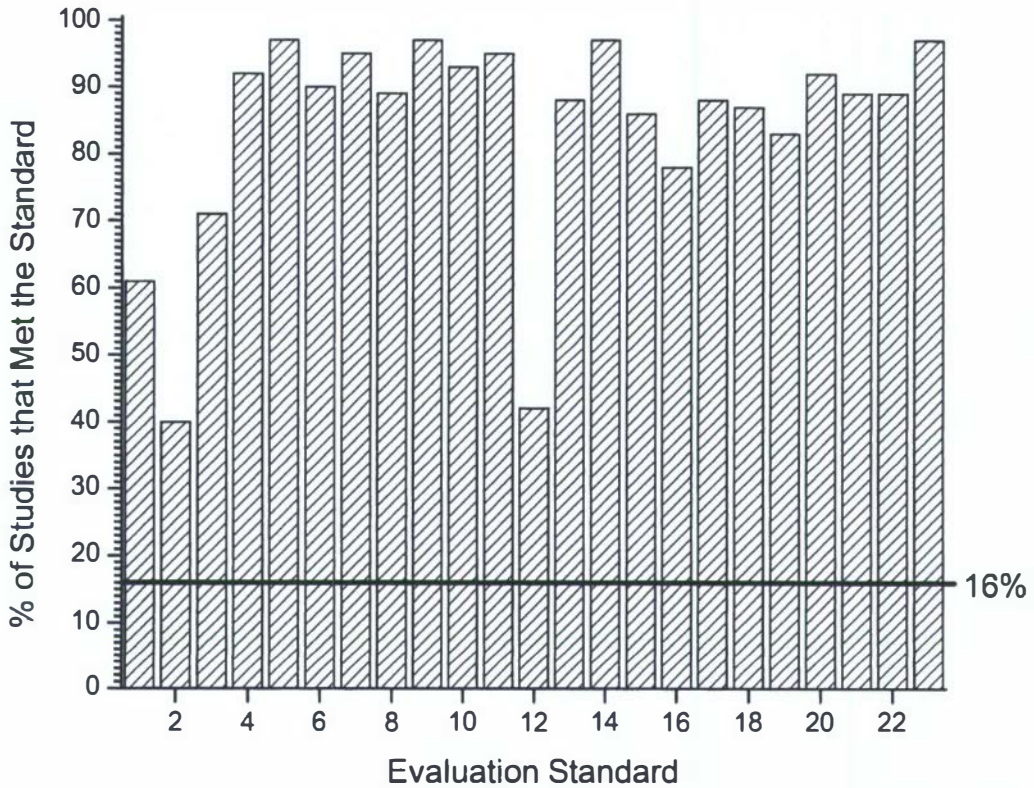


Figure 5. Standards ratings across all journals. The black line represents the total percentage of “not acceptable” ratings.

the standards for evaluating this type of research methodology as provided by Horner et al. (2005). Across seven journals and ten years, studies meeting the predefined acceptance criteria ($N = 160$) were found to have characteristics that: (a) were consistent with previous research (e.g., low rates of reported treatment fidelity, as also found by Wheeler et al., 2006) and (b) extend previous work in this area (e.g., low rates of adequate reporting of participant characteristics).

Within the limitations of the current study's scope and definitions, few IVs identified within the literature meet the NPDC's definition for EBP for learners with ASD (i.e., three different investigators or research groups must have conducted five high quality single-subject design studies). Though authors are well diversified across all articles within the current study, the number of studies within IV categories becomes a confound (e.g., 11 out of 17 IV categories had fewer than 5 studies

across the review period), as well as an uneven distribution across the geographic origins of the research (a point of comparison highlighted within the Horner model). In addition, examinations of standards ratings across IVs and journals reveal a high variability across most EBP standards, as well as a negative pattern within several standards integrally tied to external validity (e.g., description of participants, participant selection process, description of setting, and implementation fidelity). This issue is further compounded by the low numbers of participants typically found in single-subject research, a function of the historical, philosophical, and experimental origins of this type of methodology as well as certain design limitations that make it difficult (but not impossible) to include in a study more than, for instance, three participants. (Especially in the case of ASD, the argument of poor availability of participants is quickly becoming less valid.) These findings raise substantial questions re-

Table 5

Grouped standards ratings across journals.

Standard	ETDD	ETC	FOCUS	JABA	JADD	JPBI	RDD
1	✓	○	✓	●	✓	○	⊗
2	⊗	✓		●	○	⊗	⊗
3		✓	✓	○	✓	○	○
4		✓	✓	✓	✓		✓
5	✓	✓	✓	✓	✓	✓	✓
6	✓	✓	✓	✓	✓	○	✓
7	✓	✓	✓	✓	✓	✓	✓
8		✓	✓	✓	✓	✓	✓
9	✓	✓	✓	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓	✓
11	✓	✓	✓	✓	✓	✓	✓
12		●	✓	●	○	⊗	⊗
13	✓	✓	✓	✓	✓	✓	○
14	✓	✓	✓	✓	✓	✓	○
15		✓	✓	✓	✓	✓	○
16		✓	✓	✓	✓	⊗	○
17		✓	✓	✓	✓	○	✓
18		✓	✓	✓	✓	✓	✓
19	○	✓	○	✓	✓	✓	✓
20	✓	✓	✓	✓	✓		✓
21	✓	○	✓	✓	✓		✓
22	○	✓	✓	✓	✓	✓	✓
23	✓	✓	✓	✓	✓	✓	✓

Note.

- ✓ = 80% (or higher) studies met the standard
- = 70% (down to 51%)
- ⊗ = 50% (down to 31%)
- = 30% (or below)

garding the adherence of this body of research to EBP standards. However, it does appear that adherence to the standards has been improving over time. This is illustrated in the fact that the number of single-subject research studies conducted with learners with ASD has tended to increase from year to year, yet poor compliance with EBP standards has tended to decrease, though poor compliance still seems to stubbornly persist within the aforementioned areas.

Implications for Practice

Variance exists regarding the quality of research, and therefore the stakes of using the findings of research in daily practice are quite high for all involved. This creates an impetus for practitioners to become more sophisticated consumers of research. How-

ever, as in other industries such as the medical and airline industries, it should not be the responsibility of the consumer of research to perform regular, intensive quality control of the product. Rather, it is the responsibility of the consumer to know important indicators of a potentially inadequate product in order to better safeguard health and well-being, ultimately through the exercise of one's consumer rights (e.g., to shop elsewhere or file a complaint with the appropriate regulatory agency). This implies that in addition to increasing scrutiny at the level of the consumer, we must also: (a) increase scrutiny at the researcher and peer-review levels, and (b) vary the extent and type of scrutiny at each of these levels in accordance with the roles and responsibilities of those who work within them.

Implications for Research

Studies need to improve on the identification and description of participants, especially given the variability of characteristics associated with ASD and the issue of utility/applicability of research findings for practitioners. Many studies within this review *appeared* to have utilized samples of convenience (i.e., recruiting participants who were “handy” rather than searching out and targeting specific individuals based on relevant criteria). It is not clear whether this appearance was accurate or simply due to a failure to report procedure, as is the potential issue with other quality indicators such as treatment fidelity. (Duly noted are common constraints such as limitations on manuscript length, as imposed by most print journals.)

Studies also need to improve the description of the settings in which they are conducted. In order for accurate replication of the IV to occur, relevant environmental features must be replicated. Complex environments such as “classrooms” or “schools” have to be, within reason, deconstructed for the consumer so that adequate external validity can be confirmed through (in part) appropriate, accurate application of the IV in settings with features as identical as possible to the original.

As in a growing number of other studies (e.g., Gresham, MacMillan, Frankenger, & Bocian, 2000; McIntyre, Gresham, DiGennaro, & Reed, 2007; Wheeler et al., 2006), the current study found alarmingly low levels of reporting procedures and data related to treatment fidelity. Studies need to improve the reporting of *both* procedures and data in this area, for one cannot reliably assert that a functional relationship between the dependent variable and the IV has been established if one does not adequately demonstrate that the IV was implemented as designed.

There is work to do at all levels in increasing the identification and use of EPB in the education and treatment of learners with ASD. Though some indicators examined over time show that we are getting incrementally closer to this goal, other indicators show the ongoing

need for refinement of how we identify and use EBPs for and with this population. As new educational tools and treatments become available to us, this process will be never-ending. Therefore, as those who are advocates for and positive intervention agents in the lives of people with ASD, professionals at all levels are encouraged to embrace and apply this important concept.

References

- Centers for Disease Control (2009). Prevalence of autism spectrum disorders—Autism and developmental disabilities monitoring network, United States, 2006. *Morbidity and Mortality Weekly Report*, December 18, 2009; 58 (SS-10).
- Gresham, F. M., MacMillan, D. L., Frankenger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, 15, 198–205.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis*, 40, 659–672.
- National Professional Development Center (NPDC) on Autism Spectrum Disorders (2009, January). *Evidence-based practices for children and youth with autism spectrum disorders*. Retrieved from http://www.fpg.unc.edu/~autismPDC/resources/resources_public_ebp.cfm
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, 2, 137–148.
- Simpson, R. L. (2005). Evidence-based practices and students with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, 20, 140–149.
- Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities*, 21, 45–54.