Lessons Learned: A Pilot Study on Occupational Therapy Effectiveness for Children With Sensory Modulation Disorder

Kennedy Jane Miller, Sarah A. Schoen, Katherine James, Roseann C. Schaaf

KEY WORDS
• effectiveness
• occupational therapy
• pediatric
• sensory integration
• sensory modulation disorder (SMD)
• sensory processing

OBJECTIVE. The purpose of this pilot study was to prepare for a randomized controlled study of the effectiveness of occupational therapy using a sensory integration approach (OT-SI) with children who have sensory processing disorders (SPD).

METHOD. A one-group pretest, posttest design with 30 children was completed with a subset of children with SPD, those with sensory modulation disorder.

RESULTS. Lessons learned relate to (a) identifying a homogeneous sample with quantifiable inclusion criteria, (b) developing an intervention manual for study replication and a fidelity to treatment measure, (c) determining which outcomes are sensitive to change and relate to parents' priorities, and (d) clarifying rigorous methodologies (e.g., blinded examiners, randomization, power).

CONCLUSION. A comprehensive program of research is needed, including multiple pilot studies to develop enough knowledge that high-quality effectiveness research in occupational therapy can be completed. Previous effectiveness studies in OT-SI have been single projects not based on a unified long-term program of research.

All previous studies regarding OT-SI effectiveness asked a naïve question, “Is OT-SI effective?” A single study cannot answer that multifaceted question. No study described a series of prior pilot studies or met all of the criteria for a reliable and valid outcome study. Thus, the only conclusion that can be drawn after 35 years of single, non-programmatic research projects is that the evidence neither disproves nor confirms that OT-SI is effective.

The “gold standard” effectiveness study is a randomized controlled trial (RCT) in which the targeted treatment is compared to another treatment or treatments, for example, an active placebo or a passive placebo (Boruch, 1997). A rigorous RCT requires the following:

1. A homogeneous sample identified with replicable criteria
2. A “manualized” intervention approach (based on a manual detailing the intervention) and a fidelity to treatment measure that evaluates whether the intervention is completed as intended
3. Outcome measures that are sensitive enough to detect changes during the treatment duration and meaningful enough to measure the changes that the treatment purports to effect
4. Rigorous research design and methodology (e.g., blinded examiners, random assignment)

The following discussion details each of the four criteria so that future occupational therapy studies can adhere to these principles.

**Homogeneous sample.** A homogeneous sample must be narrowly defined using replicable (quantitative) measures. Previous OT-SI studies included heterogeneous (broad) samples such as combinations of children and adults with mental retardation (Close, Carpenter, & Cibiri, 1986), learning disabilities (Carte, Morrison, Sublett, Uemura, & Setrakian, 1984; Werry, Scaletti, & Mills, 1990), and aphasia (DePauw, 1978) and individuals with “at-risk” diagnoses (White, 1979). New nosologies suggest multiple subtypes of SPD criteria that were unavailable for previous studies (Interdisciplinary Council, 2005; Miller et al., 2005; Zero to Three, 2005). Study inclusion criteria should specify the subtype of SPD, as well as other factors (e.g., cognitive level, age, comorbid diagnoses).

**Replicable treatment.** None of the studies has published manuals that can be reviewed detailing the elements of the treatment. One study purports to have used a manual (Polatajko, Law, Miller, Schaffer, & Macnab, 1991), but the manual is not available for review, so study replication is not possible. The Polatajko et al. study clearly demonstrated the import of a manualized approach. A comparison of sensory integration treatment to perceptual–motor treatment found that both groups demonstrated significant changes after occupational therapy and that group differences were not significant. Because the treatments are not easily distinguished, the importance of this finding is difficult to understand.

Additionally, previous studies did not have a fidelity to treatment process. Fidelity measures provide crucial evidence evaluating adherence to treatment principles (Kazdin, 1994).

**Sensitive and meaningful outcomes.** In general, the existing studies evaluated outcomes that were neither sensitive to small increments of change nor meaningful based on parents’ priorities for treatment outcomes (Cohn, Miller, & Tickle-Degnen, 2000). The three outcomes of greatest importance to parents of children with SPD are social participation, self-regulation, and perceived competence/self-esteem and self-confidence (Cohn, 1999).

No information on the sensitivity of the outcome measures was published before previous studies’ findings. Power, therefore, was not evaluated before the study was implemented. Because adequate power was rarely achieved, Type II errors abounded in studies that concluded that OT-SI was ineffective; for example, in most cases not enough power was present to demonstrate significance of group differences, although group differences were found.

**Rigorous methodology.** Rigorous methodology for RCTs includes (a) random assignment to two or more treatment groups: experimental (i.e., occupational therapy), active treatment placebo (e.g., tutoring, special education, or play time), or passive placebo (i.e., no treatment, such as a wait-list condition); (b) evaluators blinded to group assignments; (c) appropriate research designs; and (d) adequate power. No previous studies used this methodology.

**Preparation for the Described Study**

From 1995 to 2005, a series of pilot studies was conducted before implementing an RCT of OT-SI (Ahn, Miller, Milberger, & McIntosh, 2004; Cohn et al., 2000; Mangeot et al., 2001; McIntosh, Miller, Shyu, & Dunn, 1999; McIntosh, Miller, Shyu, & Hagerman, 1999; Miller et al., 1999; Miller, Reisman, McIntosh, & Simon, 2001; Miller, Robinson, & Moulton, 2004; Miller & Summers, 2001; Miller, Wilbarger, Stackhouse, & Trunnell, 2002; Schaaf & Miller, 2005), including a wide variety of studies related to the four criteria for rigorous research studies noted previously.

**Research Questions**

This article describes a single group pretest–posttest pilot outcome study designed to inform a future RCT, specifically the following questions:
• What replicable criteria can identify a homogeneous group of children with SPD?
• How can the essential components of treatment be manualized and a fidelity to treatment measure developed?
• What outcome measures are sensitive to changes over a 20-session time period (10 weeks, twice a week)?
• What procedures can assure rigor in research design and methodology?

Method

Participants: Research Question 1

The participants were children who met global criteria for one subtype of SPD, sensory modulation disorder (SMD), based on the impression of an occupational therapy master clinician at The Denver Children’s Hospital after completion of a comprehensive occupational therapy evaluation lasting from 2 to 4 hours. The evaluation included a standardized scale, such as the Sensory Integration and Praxis Test (Ayres, 1989) for children ages 5.5–8.11 years or the Miller Assessment for Preschoolers (Miller, 1988) and FirstSTEP (Miller, 1993) for children ages 3.6–5.5 years, and clinical observations (Blanche, 2002). Because no standardized tests for SMD exist, selection was based on evaluators’ global clinical impression. Inclusion criteria also included developmental and medical history suggesting SPD and behaviors consistent with the disorder (Ayres, 1989).

Exclusion criteria were IQ < 85 and any other developmental, psychiatric, neurological, or orthopedic condition except attention deficit hyperactivity disorder (ADHD), learning disorder, and mild Tourette’s syndrome. Children could have other SPD subtypes if SMD was present. Thirty children, ages 3.9–11 years (mean age 6.79, SD = 1.75), met inclusion or exclusion criteria for SMD based on the global impression of occupational therapists.

Exclusion criteria were IQ < 85 and any other developmental, psychiatric, neurological, or orthopedic condition except attention deficit hyperactivity disorder (ADHD), learning disorder, and mild Tourette’s syndrome. Children could have other SPD subtypes if SMD was present. Thirty children, ages 3.9–11 years (mean age 6.79, SD = 1.75), met inclusion or exclusion criteria for SMD based on the global impression of occupational therapists.

Quantification of inclusion criteria included development of a parent report measure of SMD behaviors, the Short Sensory Profile (McIntosh, Miller, Shyu, & Dunn, 1999), and abnormal physiologic reactivity on electrodermal reactivity (Miller et al., 1999) (see Instrumentation). Table 1 displays demographic characteristics of the study participants.

The Experimental Treatment—OT-SI: Research Question 2

The intervention, OT-SI (Ayres, 1972; Koomar & Bundy, 2002; Parham & Mailloux, 2001), was administered twice a week for 10 weeks. Occasionally, a therapist or child missed a session due to illness, but sessions were made up within 2 weeks. The intervention was based on principles defined by Ayres (1972), emphasized use of a clinical reasoning process (Mattingly, 1991), and focused on attaining occupational goals. Therapists met bimonthly to review treatment videotapes and engage in a reflective process of understanding why the therapist made the clinical decisions observed during treatment. The team elucidated the elements of the therapeutic process and drafted a pilot manual, naming the process STEP–SI (Miller et al., 2002). This approach involved asking a series of questions to refine each child’s intervention. Based on the elements of intervention specified in the treatment manual, a draft fidelity measure was constructed and has recently been substantially elaborated (Parham et al., 2007).

Instrumentation: Research Question 3

The study tested numerous measures to determine which were most suitable for future studies. The domains and specific outcome measures are the following:

• Sensory Functioning. In iterative studies, content, item, and factor analyses of the Sensory Profile (Dunn, 1999) were conducted to create the Short Sensory Profile (McIntosh, Miller, Shyu, & Dunn, 1999). The 38 items on the Short Sensory Profile target only sensory behaviors, with a stable factor structure corresponding to sensory constructs hypothesized in SMD.

• Attention, Impulsivity, and Activity Level. Several tools were used to evaluate attention, including (a) the ADD–H Comprehensive Teacher’s Rating Scale (ACTeRS) (Ullmann, Sleator, & Sprague, 1991), which is effective in discriminating between children with and without attention disorders (Ullmann & Sleator, 1985, 1986; Ullmann, Sleator, & Sprague, 1984); (b) three subtests of the Leiter International Performance Scale–Revised: Parent Rating Scale (Roid & Miller, 1997), which has an excellent national standardization and impressive reliability and validity characteristics; (c) Barkley’s Behavior Rating for
AD/HD, widely cited in ADHD research (Barkley & Murphy, 1998); and (d) the SNAP–IV used in the National Institutes of Health (NIH) Multi-Site Trial (MTA Cooperative Group, 1999).

- **Anxiety.** Children with SMD are described as being overly anxious (Kinnealey, 1998; Kinnealey & Fuick, 1999; Pfeiffer & Kinnealey, 2003). The Multi-Dimensional Anxiety Scale for Children (MASC; March, 1997) was selected because it differentiates pathological anxiety from fears that are a natural part of development (March, 1995; Silverman, LaGreca, & Wasserstein, 1995) and distinguishes children older than age 8 years with and without anxiety disorders (March, 1997).

- **Activities of Daily Living.** The most widely used adaptive scale, the Vineland Adaptive Behavior Scale (Stinnett, Havey, & Oehler-Stinnett, 1994; Wodrich & Barry, 1991), was tested. The Vineland scale has been validated for accurate discrimination of deficits in performance of daily living skills (Altman & Mills, 1990; Douhitt, 1992; Rosenbaum, Saigal, Szatmari, & Hoult, 1995; Voelker, Shore, & Brown-More, 1999).

- **Social and Emotional Behaviors.** The Child Behavior Checklist (CBCL; Achenbach, 1991) measures social and emotional behaviors based on parent and teacher reports and was piloted because numerous critiques have substantiated its use in research (Elliott & Busse, 1992; Mooney, 1984). The construct, content, and criterion validity of the CBCL for discriminating social and behavioral issues is well established (Chen, Faraone, Biederman, & Tsuang, 1994; Jensen, Wantanabe, Richters, & Roper, 1996; Macmann, Barnett, Berd, & Jones, 1992).

- **Physiologic Measures.** Extensive evaluation of physiologic measures of sensory reactivity was completed before this study (Mangeot et al., 2001; McIntosh, Miller, Shyu, & Hagerman, 1999; Miller et al., 1999, 2001) to ascertain quantitative physiological markers assisting to define a homogeneous sample. The physiologic outcome measure used was electrodertal reactivity (EDR). EDR measures changes in the electrical conductance of the skin and is a marker of activity in the sympathetic nervous system. The children enter a pretend “spaceship” and watch the movie Apollo 13 while electrodes are attached to their hands. Data are continuously collected during the Sensory Challenge Protocol (Miller et al., 1999), a 15–20 min protocol in which 50 sensory stimuli are administered (10 stimuli in each of five sensory systems).

- **Changes in Natural Settings.** The children were videotaped for approximately 30 min before and after intervention in playtime and dinnertime. Transcripts were made of the videotaped activities, and a coding scheme was developed following the procedures outlined in Lofland and Loiland (1995). Behavior codes were verbal interactions, nonverbal socialization, self-initiation, challenges encountered, success in resolution, and type of sensory input. Five-min segments were randomly selected from five 30-min tapes, which two investigators coded independently. The total number of behaviors in each category was summed before and after intervention (intrarater reliability = .90) (Schaaf et al., 2001).

- **Individualized Measure of Parent-Perceived Priorities for Change.** A contextually relevant measure of change, goal attainment scaling (GAS; Kiresuk, Smith, & Cardillo, 1994), was explored as an outcome. GAS examines individual priorities for change that are not represented by items in standardized scales because they are idiosyncratic to the child. GAS is gaining recognition as a valid outcome of individualized change over time (Kiresuk et al., 1994). Researchers are discussing the advantages of evaluating outcomes that are patient-centered and specific for each participant in a study (Abikoff, 2001; Rockwood et al., 2000). Although the goals are different for each participant, the score is standardized by writing goals that have responses that are theoretically spaced the same distance apart (e.g., the same level of difficulty to achieve) (Forbes, 1998). Thus, a mathematical method can be derived of calculating the extent to which the goals are met (Kiresuk & Sherman, 1968). Table 2 displays an example of a GAS goal.

Recent studies have demonstrated the use of GAS in outcome studies for people with a wide variety of disabilities: lower-extremity amputations (Rushton & Miller, 2002), traumatic brain injury (Joyce, Rockwood, & Mate-Kole, 1994; Malec, 2001), cognitive rehabilitation (Rockwood, Joyce, & Stolee, 1997), motor delays in infants (Palisano, 1993), and geriatrics (Stolee, Rockwood, Fox, & Streiner, 1992). In some studies, GAS has been found to be more responsive to intervention than norm-referenced standardized measures (Rushton & Miller, 2002). Interrater reliability has been found to be moderate to excellent (.67, Joyce et al., 1994; .67, Rushton & Miller, 2002). Current

<table>
<thead>
<tr>
<th>Rank</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>–2</td>
<td>Child does not interact with peers.</td>
</tr>
<tr>
<td>–1</td>
<td>Child always plays alone during recess. Does not interact with peers during recess.</td>
</tr>
<tr>
<td>0</td>
<td>Child interacts with 1 peer with structure and/or cues and assistance.</td>
</tr>
<tr>
<td>+1</td>
<td>Child interacts with 1–2 peers independently on a consistent basis.</td>
</tr>
<tr>
<td>+2</td>
<td>Child interacts with a small group of peers independently.</td>
</tr>
</tbody>
</table>

Table 2. Sample Goal Attainment Scale Items

*Note. Using Ottenbacher and Cusick’s (1993) method first, the expected performance (0) is defined. Other possible outcomes then are established: less than expected level (–1), much less than expected level (–2), greater than expected level (+1), and much greater than expected level (+2). Parents rank goals by priority and degree of difficulty.*
studies will further refine this tool for the study of the effectiveness of OT-SI (Mailloux et al., 2006).

**Procedures: Research Question 4**

After informed consent was obtained from parents, the parent report scales and Vineland interview were completed in a private, 1-hr, semistructured interview. A trained occupational therapist viewed a videotape of the parent interview to complete the writing of the GAS.

**Results**

Because this was a pilot study, the following findings highlight process issues relating to the four required principles of an RCT rather than focusing on quantitative outcome data: (a) inclusion criteria, (b) a manualized approach and fidelity to treatment, (c) sensitive and appropriate outcome measures, and (d) procedures for rigorous methodology.

**Inclusion Criteria**

Inclusion was based on the master clinician’s global impression of SMD. The quantitative use of both the Short Sensory Profile (McIntosh, Miller, Shyu, & Dunn, 1999) and EDR also was piloted as inclusion criteria with promising but not definitive results. Further study of both measures is indicated to refine cut points for subject inclusion in future studies.

**Replicable Treatment**

A preliminary intervention manual was completed and published for use in future studies (Miller et al., 2002). A fidelity rating scale was piloted. Additional work is indicated to refine both measures.

**Outcome Measures**

A large number of outcome variables were studied to determine which were most sensitive to change. Paired t-test statistics were used and the findings are summarized in Table 3 for each measure that was significant or showed a trend in the hypothesized direction.

The mean change over the 20-session period with significance values and effect sizes are noted in Table 3. Measures that were nonsignificant or did not show changes in the hypothesized direction included Barkley’s Behavior Rating Scale (Barkley & Murphy, 1998), the SNAP–IV (MTA Cooperative Group, 1999), the MASC (March, 1997), three subtests of the Vineland (Stinnett et al., 1994), subtests of the CBCL (Achenbach, 1991), subtests of the Leiter–R (Roid & Miller, 1997) except Attention, some EDR variables, and videotaped changes in natural settings based on the current paradigm. Measures considered sensitive to change had effect sizes that ranged from 0.29 to 2.16. As a pilot study of 30 children, outcomes with effect sizes larger than 0.53 could be detected with 80% power (Type I error rate = .05). The measures in Table 3 warrant use in future randomized clinical pilot studies. Conclusions about reliability of which measures demonstrated reliable changes are tentative due to multiple comparisons (e.g., the probability exists that some significant differences occurred by chance).

**Rigorous Methodology**

Myriad unexpected challenges occurred; for example, the long wait time for Internal Review Board approval, the number of canceled appointments due to illness and vacations, the maintenance of a complex database, and staff turnover. In short, this study took much longer than anticipated; however, it provided extensive information about methodological sources of error to control in future RCT studies.

**Discussion**

This study highlights the need for pilot studies to inform the process of RCTs before implementation. The design and implementation of rigorous research require time,

---

**Table 3. Results of Single Group Pilot Study**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Leiter–R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td>24</td>
<td>5.88</td>
<td>1.94</td>
<td>22</td>
</tr>
<tr>
<td>Cognitive/Social</td>
<td>24</td>
<td>76.83</td>
<td>10.59</td>
<td>22</td>
</tr>
<tr>
<td>SSP Total Score</td>
<td>30</td>
<td>–3.39</td>
<td>1.92</td>
<td>27</td>
</tr>
<tr>
<td>Vineland Socialization</td>
<td>24</td>
<td>79.04</td>
<td>12.66</td>
<td>19</td>
</tr>
<tr>
<td>CBCL Externalizing</td>
<td>30</td>
<td>60.93</td>
<td>9.79</td>
<td>21</td>
</tr>
<tr>
<td>Internalizing</td>
<td>30</td>
<td>61.57</td>
<td>10.51</td>
<td>21</td>
</tr>
<tr>
<td>GAS</td>
<td>27</td>
<td>30.37</td>
<td>1.17</td>
<td>27</td>
</tr>
</tbody>
</table>

*Note.* Leiter–R = Leiter International Performance Scale–Revised: Parent Rating Scale (Roid & Miller, 1997); SSP = Short Sensory Profile (McIntosh, Miller, Shyu, & Dunn, 1999); Vineland = Vineland Adaptive Behavior Scale (Stinnett, Havey, & Dehler-Stinnett, 1994); CBCL = Child Behavior Checklist (Achenbach, 1991); GAS = goal attainment scaling (Kiresuk, Smith, & Cardillo, 1994).
stamina, and patience. Outcome research is a long-term process, not a single project, as is typical of previous literature. The contradictory results in the previous 80 studies are likely due to a lack of pilot research. In summary, the following crucial lessons were learned relative to the four essential principles for conducting rigorous RCTs.

**Defining a Homogeneous Group**

Inclusion criteria identifying a homogeneous group of children is crucial. A combination of a behavioral measure and a physiologic measure appears useful. However, further research is needed to determine cut points for participant inclusion.

**Establishing a Manualized Approach to Intervention**

The intervention methodology was refined and published (Miller et al., 2002). With subsequent NIH funding, researchers from five sites around the United States continue to work on the manualized approach for OT-SI treatment for future effectiveness research. A study of the reliability and validity of the fidelity to treatment measure is forthcoming (Parham et al., 2007).

**Choosing Sensitive and Appropriate Outcomes**

Many outcome measures were evaluated to determine sensitivity in detecting improvement over 10 weeks. The following measures showed the most potential for future study: the total test score of the Short Sensory Profile (McIntosh, Miller, Shyu, & Dunn, 1999), the GAS (Kiresuk et al., 1994), the Socialization subtest of the Vineland Adaptive Behavior Scale (Stinnett et al., 1994), the composite scales of the CBCL (Achenbach, 1991), and the Attention subtests and Cognitive/Social Composite of the Leiter International Performance Scale–Revised: Parent Rating Scale (Roid & Miller, 1997). Identifying specific outcome measures from the large array field tested decreases spurious findings for future studies.

The GAS showed the most prepost change during the pilot study, was well received by parents, and captured the outcomes of most import to families (Cohn, 1999). Additional research has been undertaken to study the reliability and validity of the GAS with this population (Mailloux et al., 2007).

**Using Rigorous Methodology**

Procedures that can be implemented in the next study to assure increased rigor in research design and methodology were defined. Of utmost importance is randomization to an active or a passive treatment control.

This article shows the importance of conducting pilot effectiveness studies. Using knowledge gleaned from this pilot study, the authors plan to conduct a pilot RCT before conducting a large-scale RCT study. Researchers interested in occupational therapy effectiveness research should implement pilot studies that build on each other instead of attempting to do the “one perfect” study. Only by building programmatic research can a body of knowledge be created to move the empirical basis of the profession forward. ▲

**Note**

1In a later iteration, the first author modified the name STEP–SI to “A SECRET” (Miller, 2006). Using “A SECRET,” the parent is taught the therapy “secrets” that regulate the specific child, increase his or her social participation and self-confidence or self-esteem, and then address other specific occupational goals of the family.

**Acknowledgments**

First and foremost, the authors thank the children and their parents who participated in this study. Without the gifted therapists at The Children's Hospital of Denver, who implemented the occupational therapy, this study would not have been completed: Julie Butler, Becky Greer, Corrine Jack, Nicki Pine, Robin Seger, Tracy Stackhouse, Clare Summers, Sharon Trunnel, and Julie Wilbarger. The KID Foundation and the STAR Center staff, who worked tirelessly over 10 years to obtain the preliminary information needed to complete this project, made invaluable and priceless personal sacrifices and major contributions, particularly Jude McGrath. In addition, we thank the graduate school faculty and students of Thomas Jefferson University, Occupational Therapy Department, for their efforts related to coding the videotapes. We also thank Barbara Brett-Green, Janice Burke, Ellen Cohn, Jane Kohmar, Zoe Mailloux, Teresa May-Benson, Diane Parham, Susanne Smith Roley, and Clare Summers for contributing ideas about goal attainment scaling and the OT-SI treatment protocol and fidelity measure. Primary funding for this work was provided by an NIH Mentored Research Scientist Development Award (1K01HD001183-01A1), the Wallace Research Foundation, and the American Occupational Therapy Foundation. Additional support was provided by an R21 NIH One-Year Planning grant (#1 R21 HD41614-01), the General Clinical Research Centers Program at The Children's Hospital (#M01 RR00069), The Children's Hospital Research Institute Scholar Award, and the Coleman Institute for Cognitive Disabilities research fellowship.

**References**


March/April 2007, Volume 61, Number 2


