A Meta-Analysis of Research on Sensory Integration Treatment

Sadako Vargas, Gregory Camilli

Key Words: outcome studies • quantitative method

Objective. The purpose of this study was to find whether existing studies of treatment using sensory integration approaches support the efficacy of these approaches.

Method. With meta-analysis, the results of sensory integration efficacy research studies published from 1972 to the present were synthesized and analyzed. Sixteen studies were used to compare sensory integration effect with no treatment (SI/NT), and 16 were used to compare sensory integration effect with alternative treatments (SI/ALT). Overall average effect sizes, comparisons of the effect sizes for different dependent variables, and secondary factors associated with effect size variation were examined.

Results. The weighted average effect size of SI/NT studies was .29. However, there was a significant difference between the average effect sizes of the earlier studies (.60) and the more recent studies (.03). Of the outcome measures, larger effect sizes were found in the psychoeducational category (.39) and motor category (.40). Of SI/ALT studies, the average effect size was .09, not significantly different from zero.

Conclusion. Three central conclusions can be made. First, in the SI/NT comparison, a significant effect was replicated for sensory integration treatment effects in earlier studies, but more recent studies did not show overall positive effects. Second, larger effect sizes were found in psychoeducational and motor categories. Third, sensory integration treatment methods were found to be as effective as various alternative treatment methods.


Treatment using sensory integration approaches has gained popularity among parents as an effective approach to remediation, especially in pediatrics. However, the concept of sensory integration has been the subject of criticism and controversy in the fields of neuropsychology, education, and medicine (Committee on Children With Disabilities, 1985; Hoehn & Baumeister, 1994; Lerer, 1981). Some professionals support and appreciate the value of sensory integration approaches, whereas others consider them to be “demonstratively ineffective” methods of intervention (Hoehn & Baumeister, 1994). Treatment sessions are costly and often last more than 1 year. Thus, an evaluation of the efficacy of sensory integration treatment approaches is important from ethical, financial, and rehabilitative points of view. The major topics that need to be examined are the efficacy of sensory integration treatment approaches (a) compared with other intervention strategies, (b) in different behavioral domains, and (c) for different diagnostic and age groups. This article reviews literature on efficacy studies of sensory integration treatment approaches and provides a meta-analysis of studies from 1972 to 1994.
Literature Review

Of the five literature reviews of efficacy of sensory integration treatment approaches that have previously appeared, only Ottenbacher's (1982) was a meta-analysis. In the remaining four studies—Arendt, MacLean, and Baumeister (1988), Hoehn and Baumeister (1994), Polatajko, Kaplan, and Wilson (1992), and Schaffer (1984)—the reviewers conducted traditional literature reviews, deriving their conclusions from logical and critical analyses of published efficacy studies. Ottenbacher's review examined eight experimental studies published between 1972 and 1981 with respect to sensory integration efficacy. Each study was coded in terms of (a) subject type; (b) design characteristics, including subject selection and assignment, total hours of treatment, and dependent variables; (c) statistical outcomes; and (d) publication location, year, and source. The eight studies reported 47 outcome measures, each with an associated statistical test. The mean effect size was .79 across all outcome measures, p < .001. These outcomes were then broken down separately for dependent variable types (motor or reflex performance, academic performance, language function) and subject type (mental retardation, learning disability, aphasia, at risk).

For the dependent variable classes, the mean effect size was 1.03 for the motor–reflex variables, .75 for the academic performance variables, and .43 for the language variables. Comparing the sensory integration effect among the diagnostic groups, the mean effect sizes were .52 for the subjects with mental retardation, .68 for the subjects with learning disabilities, and 1.20 for subjects with aphasia or who were at risk. On the basis of these findings, Ottenbacher (1982) concluded that sensory integration treatment approaches had a positive impact when applied to the populations investigated. However, Hoehn and Baumeister (1994) criticized Ottenbacher's meta-analysis for both the inadequacies of the original (primary) studies and for the method of analysis. The latter criticism focused on the problem of data nonindependence due to the use of outcome as the unit of analysis and the exclusion of studies with alternative treatment control groups. To date, no study has attempted to replicate Ottenbacher's meta-analysis.

Of the four traditional literature reviews, three examined efficacy studies of sensory integration treatment approaches with subjects with learning disabilities (Hoehn & Baumeister, 1994; Polatajko et al., 1992; Schaffer, 1984), and one examined efficacy with subjects with mental retardation (Arendt et al., 1988). All concluded that there was insufficient evidence to support the efficacy of sensory integration treatment approaches but differed in their analytic approaches and rationales for the conclusions. Supportive evidence for sensory integration treatment approaches was substantially weakened by inadequate experimental design.

Schaffer (1984) reviewed five early studies by Ayres published between 1972 and 1976 that Ottenbacher (1982) included in his meta-analysis and identified a number of threats to internal validity. Schaffer listed the methodological flaws as (a) inadequate sampling and matching procedures, (b) lack of definitions of the subject population, (c) potential Hawthorn effects, (d) erroneous use of gain scores, (e) potential tester bias, and (f) absence of information regarding reliability and validity of measures. A similar claim was made by Arendt et al. (1988), who reviewed eight studies published between 1977 and 1983 on the efficacy of sensory integration treatment approaches with subjects with mental retardation. Though all but one of these studies reported positive gains, the authors found that the studies (a) did not use blinding procedures for observers, (b) lacked information on interobserver agreement, and (c) did not control for subjects' heterogeneity.

Polatajko et al. (1992) reviewed seven studies published after 1979 pertaining to academic performance of children with learning disabilities. They examined effect sizes of the variables that were most likely to have reasonable effects, with p values between .05 and .20, and found the sensory integration effect sizes to be small. The post hoc power analysis with effect sizes of this magnitude suggested that larger samples than those used by these studies would have been required to attain significance. Polatajko et al. also found that although multiple hypothesis tests were performed in these studies, the impact of sensory integration treatment approaches achieved significance in only a few instances. Of nine hypothesis tests having a reading-related outcome, only one positive result was obtained that favored the sensory integration treatment group. For mathematical ability, only one of six hypothesis tests achieved significance at the .05 level, favoring the sensory integration treatment group. These authors also questioned the validity and sensitivity of the measures used to assess treatment effects.

Hoehn and Baumeister (1994) examined seven studies published after 1981 of treatment using sensory integration approaches with children with learning disabilities. They examined experimental effect in four categories: (a) postrotary nystagmus; (b) sensorimotor, perceptual, and motor measures; (c) cognitive, language, and academic measures; and (d) self-esteem or self-concept, attention, and behavioral measures but found only a few instances of notable effects for sensory integration therapy. This persistent criticism of sensory integration treatment approaches warrants a quantitative analysis of all the sensory integration efficacy studies published up to the present (i.e., 1995) to help resolve questions of efficacy.

Meta-Analysis Review Procedure

Literature Search

A literature search was conducted for studies published...
from 1972 (when the sensory integration theory was introduced) through 1994 with computer search systems (ERIC, PsychoLit, MEDLINE, Dissertation Abstracts), manual searches through article bibliographies, personal contact with authors, and consultation with the Sensory Integration International Association. The key terms used for the search were sensory integration, sensory integration therapy, sensory stimulation, vestibular stimulation, praxis, motor planning, learning disability, mental retardation, developmental delay, aphasia, speech and language disorder, and mental disorder. A total of 76 published articles and 5 masters’ theses were found that reported results of sensory integration efficacy.

The collected studies were subjected to the following inclusion criteria: that they (a) investigated the effect of treatment using sensory integration approaches; (b) reported comparisons of at least two conditions; (c) reported findings and results in a manner that allowed quantitative analysis; and (d) reported the outcome measures in the broad categories of academic skills, motor function, behavior, language function, and sensorimotor function. Sensory integration treatment was operationally defined as a treatment that aimed to enhance development of basic sensory integration processes with activities that provide vestibular, proprioceptive, tactile, or other somatosensory inputs as modalities to elicit adaptive body responses. The activities were typically client directed, with active participation of the person being treated. Case studies, single-group design results, and laboratory-type stimulation studies were not included in the analysis. The variations of implementation and quality of the treatment using sensory integration approaches were coded, and their effects were investigated. Twenty-two articles by 19 authors met the selection criteria.

**Definition of a Study**

The unit of analysis was the primary study, which was defined as an experiment with one treatment and one comparison group. However, several other conventions were used. First, an experiment was counted as two studies when the performance of one experimental group was compared with two different comparison groups (e.g., sensory integration and alternative treatment); however, sensory integration effect with no treatment (SI/NT) and sensory integration effect with alternative treatments (SI/ALT) studies were analyzed separately. Second, when two different articles presented different outcome measures of the same experiment, the results were combined into one study (e.g., Humphries, Snider, & McDougall, 1993; Humphries, Wright, Snider, & McDougall, 1992).

The literature selection resulted in 14 articles contributing 16 SI/NT comparison studies and 11 articles contributing 16 SI/ALT comparison studies (see Table 1). The sample sizes for each group ranged from 6 to 78 in the SI/NT studies and 5 to 35 in the SI/ALT studies. For the SI/NT studies, the total numbers of participants were 341 in the experimental groups and 237 in the control groups, studies, and laboratory-type stimulation studies were not included in the analysis. The variations of implementation and quality of the treatment using sensory integration approaches were coded, and their effects were investigated. Twenty-two articles by 19 authors met the selection criteria.

**Table 1**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Diagnosis</th>
<th>Comparison</th>
<th>Dependent Variable Category</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>P</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>L</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mental retardation (adult)</td>
<td>SI/ALT/ALT</td>
<td>L</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Minor brain dysfunction</td>
<td>SI/NT</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>P, S</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>P, S</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Mental retardation (adult)</td>
<td>SI/NT</td>
<td>B, L, M, S</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT/ALT</td>
<td>L, M, P, S</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Aphasia</td>
<td>SI/NT/ALT</td>
<td>M, S</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>At risk</td>
<td>SI/NT</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mental retardation (adult)</td>
<td>SI/NT</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT/ALT</td>
<td>L, M, P, S</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT/ALT</td>
<td>B, L, M, P, S</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Motor delay</td>
<td>SI/ALT</td>
<td>M</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Motor delay</td>
<td>SI/ALT</td>
<td>M</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mental retardation (child)</td>
<td>SI/ALT/ALT</td>
<td>B, L, M</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mental retardation (child)</td>
<td>SI/ALT/ALT</td>
<td>B, L, M</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/ALT</td>
<td>B, M, P</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/ALT/ALT</td>
<td>L, S</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>B, L, M, P, S</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/NT</td>
<td>P</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Learning disability</td>
<td>SI/ALT</td>
<td>B, M, P</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>SI/NT</td>
<td>M, P</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 234. Subscripts (1, 2) after the reference indicate that two studies were reported in one article. SI = sensory integration; NT = no treatment; ALT = alternative treatment; B = behavior; L = language; M = motor; P = psychoeducational; S = sensory-perceptual.
whereas the corresponding sample sizes in SI/ALT comparisons were 250 in the experimental groups and 191 in the control groups. The female–male ratio was approximately 1 to 6 for children and about 1 to 1 for adults. Differential scores by gender were not reported in the studies.

**Measurement Variables**

A total of 264 outcome measures was reported in the 24 primary studies for which 228 effect sizes were calculated: 122 effect sizes for the SI/NT analysis and 106 for the SI/ALT analysis. To minimize the number of the dependent variables, the outcome measures were classified into five categories: psychoeducational, behavior, language, motor, and sensory-perceptual (see Table 2). To help reduce subjective bias, the classification task was submitted to five expert judges who were three occupational therapists, one learning consultant, and one child psychologist. Judges had an initial agreement rate of about 95% before discussion and reclassification. Approximately 67% of the effect sizes were classified as either psychoeducational or sensory-perceptual in the SI/NT studies. In the SI/ALT studies, 44% of the total effect sizes were in the psychoeducational and sensory-perceptual categories.

**Design Variables**

After the studies were selected, their various features were coded or given quantitative descriptions. With the coded results, the study characteristics were categorized into 11 variables to investigate associations with the magnitude of sensory integration effects. **Treatment-related** variables were quality of treatment using sensory integration approaches, total treatment hours, diagnosis and age and children's age. **Design-related** variables were design quality, sampling method, number of outcome measures, number of measurement categories, professional affiliation of the researchers geographic location, and publication year. For two reasons, all independent variables were dichotomized, including age, publication year, sensory integration and design quality, number of outcome measures, and number of measurement categories. First, on inspection of correlations among original and dichotomized independent variables, negligible differences were found. Second, this dichotomization enabled parallel analysis across all independent variables, especially in regard to obtaining and interpreting Q statistics.

**Quality of treatment using sensory integration approaches.**

The treatment methods were first coded according to the criteria for using sensory integration approaches in treatment established by Ayres and her colleagues (Kimball, 1988), and the list of sensory integration treatment quality indicators was constructed by collapsing the multiple criteria into one scale. Points were assigned to each treatment method according to the sensory integration treatment quality indicators. The points assigned ranged from 3 (low quality) to 11 (high quality). High sensory integration quality indicated that the treatment adhered closely to the established treatment criteria, and low sensory integration quality meant that the treatment deviated from the established criteria. The average sensory integration quality scores were 7.56 for the SI/NT studies and 7.31 for the SI/ALT studies, indicating that the overall sensory integration quality was similar. For the moderator effect analysis, the sensory integration quality of treatment variable was dichotomized into low (3–7 points) and high (8–11 points).

**Total treatment hours.** Treatment hours were defined as the total hours of treatment provided to the experimental group, and this variable was calculated by multiplying the minutes per session by weekly frequency by total weeks of treatment. Two of the SI/NT studies did not report the weekly frequency or the session length, resulting in 14 studies for this analysis. The treatment hours ranged from 13 (25-min session twice weekly for 4 months) to 180 (45-min session five times weekly for 12 months). The average hours of SI/NT studies was about 60, and the average hours of SI/ALT studies was about 36. The hour variable was dichotomized into Level 1 (< 60 treatment hours) and Level 2 (> 60 treatment hours).

**Diagnosis and age.** Although six diagnostic groups were represented in the studies, examination of sensory integration effect for each was not possible because of the sparse representation of some of the diagnoses. The majority of the studies were of children with learning disabilities, and other diagnostic groups were examined in only a few studies. Additionally, most of the participants in these latter diagnostic groups were adults, leading to the confounding of age and diagnosis. Therefore, the diagnostic and age variables were combined into one, diagnosis and age, with Level 1 for the adults (> 21 years of age) with chronic conditions, such as mental retardation and psychiatric disorder, and Level 2 for the children (< 12 years of age) with learning disabilities and other developmental disorders.

**Children's ages.** The average ages of the children ranged from 4 to 10 years in SI/NT studies and 3 to 9 years in SI/ALT studies. The average age was dichotomized as Level 1 (younger group) and Level 2 (older group). The cut-off
Design quality. In previous literature reviews, methodo-
logical inadequacy was discussed as a factor that weakened
the study conclusions. Therefore, design quality, defined as
the degree of internal validity, was coded on the basis of
Cook and Campbell's (1979) formulation. Points were
assigned to individual studies on a scale of 0 to 8, with 0
indicating the best quality design. The average design qual-
ity points was 2.50 for the SI/NT studies and 2.13 for
SI/ALT studies. The design quality variable was dichoto-
mized into Level 1 (0–2) and Level 2 (3–8).

Sampling method. The sampling method variable was
dichotomized into Level 1 (broad sampling), consisting of
the studies with participants from more than one facility or
school, and Level 2 (narrow sampling), consisting of the
participants referred to clinics or residing or attending one
rehabilitation facility. The sampling method defines the
target population to whom the result could be generalized,
representing an aspect of external validity.

Number of outcome measures. This variable was defined
as the number of the effect sizes each study contributed to
the analysis. The number of the outcome measures ranged
from 1 to 29 for both SI/NT and SI/ALT comparisons.
One possible implication of the number of the outcome
measures is that using multiple measures could result in a
more reliable index of the ability being measured than
using a single measure. The average number of the out-
come measures was 7.6 for the SI/NT studies and 6.6 for
the SI/ALT studies. This variable was dichotomized into
Level 1 (1–4 outcomes) and Level 2 (> 4 outcomes).

Number of measurement categories. The outcome mea-
sures were classified into five measurement categories: psy-
chological, behavior, language, motor, and sensory-
perceptual. The average number of measurement categories
was 2.3 for SI/NT studies and 2.4 for SI/ALT studies. This
variable was dichotomized into Level 1 (1 category) and
Level 2 (more than 1 category).

Professional affiliation of the researchers. Sensory inte-
gration theory and treatment approaches are most popular
among occupational therapists. It was of interest, therefore,
to compare the results of the studies conducted by occupa-
tional therapists with those conducted by other profession-
als to examine the moderating effect of professional affili-
ation. Level 1 studies were those in which at least one
occupational therapist was included as a main author, and
Level 2 studies were conducted by other professionals. The
majority of the studies (75%) were conducted by occupa-
tional therapists.

Geographic location. The geographic location of a study
could influence the treatment effect because of various rea-
sons, such as a difference in the social and educational
backgrounds of the participants and the therapists. Within
the United States, more studies were conducted on the
West Coast where Ayres developed the sensory integration
theory and treatment approaches. The foreign countries
represented were Canada, Australia, and New Zealand,
with the largest number of the foreign studies from
Canada. The studies were classified into Level 1 (United
States) and Level 2 (foreign).

Publication year. The year variable was dichotomized
into Level 1 (1972–1982) and Level 2 (1983–1994), with
eight studies in each level. The earlier studies include those
by Ayres (1972, 1977). An examination of the sensory inte-
gration effects over the years might give some insight into
trends in efficacy.

Method of Analysis

Effect Size Calculation

The effect size was used to gauge the difference between the
pretest and posttest change scores of the treatment and
comparison groups. Effect sizes were calculated from the
pretest and posttest means and standard deviations, chi-
square statistics, gain scores extracted from graphs, and t or
F statistics with the formulas presented by Hedges and
Olkin (1985) and Cooper (1989). In this study, a positive
effect size indicated that sensory integration group per-
formance exceeded that of the comparison group, and a
negative effect size indicated that the comparison group
members showed more improvement than the sensory
integration group. An effect size of 1.0 is typically inter-
preted as the average participant in the treatment group
exceeding 84% of the participants in the control group on
the outcome measure. The value of effect size was multi-
plied by a correction factor for sample size bias to obtain an
unbiased estimator (Hedges & Olkin, 1985).

Weighing Effect Sizes

To increase the precision and reliability of the average
effect sizes (within a category of an independent variable),
a single composite weight was constructed by multiplying
three individual weights. These individual weights repre-
sent three different aspects of study design. The variance
weight is inversely proportional to the variance of each
effect size and was formulated by Hedges and Olkin
(1985); it generally provides larger weights to the studies
with larger samples. The measurement weight is 1/k where k
is the number of the effect sizes used in one study. The
measurement weights for effect sizes within a study sum to
unity, giving each study an equal contribution to analytic
results (regardless of the number of the effect sizes derived
from the study). The study weight was 1/n where n was the
number of the comparisons made with the same experi-
mental group. This weighing scheme ensures that studies
are weighted equally with regard to their impact on ana-
lytic results and gives greater weight to effect sizes that
have greater precision (i.e., smaller standard errors).
Homogeneity Analysis and Moderator Effect

Hedges’ homogeneity statistic $Q$ was used to investigate within and between variances among the levels of each independent variable (Hedges & Olkin, 1985). The overall $Q$ statistic tests the null hypothesis of equal effect sizes. If the null hypothesis is not rejected, the average effect size is considered representative of the whole group. When the hypothesis is rejected (at a particular alpha level), the alternative hypothesis that the effect sizes differ is supported, and the moderator effects contributing the difference are investigated. The overall $Q$ statistic is labeled $Q_t$, and $Q_t = Q_b + Q_w$, where $Q_b$ represents differences between levels of a moderator variable and $Q_w$ represents differences within levels of the moderator variable. The pattern of moderator variable confounding was examined with hierarchical cluster analysis.

Results of the Sensory Integration Versus No-Treatment Analysis

The weighted average effect size for all SI/NT studies was .29, significantly different from zero at the .05 level (95% confidence interval = .12-.48). The total within variance (overall $Q_b$) was 28.10, exceeding the expected variation due to the sampling error, $\chi^2(15, 16) = 28.10, p < .05$, indicating that the effect sizes were not homogeneous and, consequently, that the average effect size was not representative of all SI/NT studies. Therefore, the moderator effects of the measurement and design variables were examined.

Difference Among Measurement Categories

The $Q_b$ among category variations was not significant, $\chi^2(4, 16) = 2.88, p > .50$, indicating that there were no significant differences among the average effect sizes of the five categories. Homogeneity was attained in each category. Effect sizes for dependent variable categories are shown in Table 3.

Differences of Effects Associated With the Design Variables

The weighted average effect sizes of each level and sublevel of the design variables are shown in Table 4. Significant moderator effects were noted for three design variables—the publication year, number of measurement categories, and number of outcome measures. Only significant effects are reported.

Publication year. There was a significant difference in the magnitude of the effect sizes reported by the early (1972–1982) studies and recent (1982–1994) studies, $p = .005$. The average effect size of the early studies was .60 (95% confidence interval = .33–.86), whereas the average effect size of the more recent studies was .03 (95% confidence interval = .20–.26). Table 5 breaks down effect sizes for measurement categories by early and recent studies.

Number of outcome measures. Studies using four or fewer total outcomes tended to have a higher average effect size than those using more than four ($p < .05$). The average effect size of the former type was .53 (95% confidence interval = .28–.78), whereas the average effect size of the latter studies was .03 (95% confidence interval = .18–.32).

Number of measurement categories. The studies that examined the sensory integration effects in only one dependent variable category yielded significantly greater magnitude of the effect sizes than studies that examined the effects in two to five categories ($p = .05$). The average effect size of the measurement category was .53 (95% confidence interval = .26–.80), whereas the average effect sizes of the latter studies was .10 (95% confidence interval = .13–.33).

Cluster analysis. The cluster analysis revealed significant confounding of the three variables cited previously. The early studies used a fewer number of the outcome measures in one or two categories, and the recent studies used many outcome measures in more than three categories. Thus, it was not possible to identify unique effect of each of the three variables. For ease of interpretation, results for the three variables will simply be referred to as “chronological trend.”

Attempts were made to explore and identify the reason for the chronological trend of the decreased sensory integration effect. The results were as follows:

1. The decrease in the sensory integration effect was not due to the increased degree of improvement made by the subjects in the no-treatment group in the recent years because the pretest and posttest differences of the no-treatment group remained statistically equal through the years.
2. The scores of the common standardized tests used in both periods also decreased in the recent studies; therefore, the decreases of the sensory integration effect sizes could not be attributed to the increased number of outcome measures used in the recent studies.
3. Decrease in the effect sizes had no notable relationship to the testers’ awareness of the experimental group affiliation of the participants or the experience level of the treatment providers.
Table 4
Average Effect Sizes of Design Variable Levels and Sublevels (SI/NT)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>Effect Size</th>
<th>Sublevel</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of treatment using sensory integration</td>
<td>1 (low)</td>
<td>.23</td>
<td>1.1 (lowest)*</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2 (high)*</td>
<td>.34*</td>
<td>2.1 (highest of low)</td>
<td>9</td>
</tr>
<tr>
<td>Total number of treatment hours</td>
<td>1 (&lt; 60 hr)*</td>
<td>.35*</td>
<td>1.1 (4 years)*</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2 (&gt; 60 hr)*</td>
<td>.11</td>
<td>1.2 (6 and 7 years)*</td>
<td>6</td>
</tr>
<tr>
<td>Diagnosis and age</td>
<td>1 (adult/chronic)*</td>
<td>.21</td>
<td>2.1 (highest of low)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2 (child/learning disability)</td>
<td>.30</td>
<td>2.2 (lowest)</td>
<td>13</td>
</tr>
<tr>
<td>Children's age</td>
<td>1 (4 up to 8 years)</td>
<td>.23</td>
<td>1.1 (4 years)*</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2 (8 to 10 years)*</td>
<td>.39*</td>
<td>1.2 (6 and 7 years)*</td>
<td>6</td>
</tr>
<tr>
<td>Design quality</td>
<td>1 (high)*</td>
<td>.36*</td>
<td>2.1 (highest of low)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2 (low)</td>
<td>.18</td>
<td>2.2 (lowest)</td>
<td>7</td>
</tr>
<tr>
<td>Sampling method</td>
<td>1 (broad sampling)*</td>
<td>.33*</td>
<td>2.1 (Australia)*</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2 (narrow sampling)*</td>
<td>.25</td>
<td>2.2 (New Zealand)*</td>
<td>8</td>
</tr>
<tr>
<td>Number of outcome measures</td>
<td>1 (few)*</td>
<td>.53*</td>
<td>2.3 (Canada)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2 (many)*</td>
<td>.07</td>
<td>. (Significantly different from zero at alpha = .05)</td>
<td>8</td>
</tr>
<tr>
<td>Number of measurement categories</td>
<td>1 (few)*</td>
<td>.53*</td>
<td>1.1 (early studies)*</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2 (many)*</td>
<td>.10</td>
<td>1.2 (recent studies)*</td>
<td>9</td>
</tr>
<tr>
<td>Geographic location of study</td>
<td>1 (United States)*</td>
<td>.25*</td>
<td>2.1 (Australia)*</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2 (foreign)</td>
<td>.33*</td>
<td>2.2 (New Zealand)*</td>
<td>8</td>
</tr>
<tr>
<td>Professional affiliation of the researchers</td>
<td>1 (occupational therapy)</td>
<td>.37*</td>
<td>1.1 (early studies)*</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>2 (not occupational therapy)*</td>
<td>.03</td>
<td>1.2 (recent studies)*</td>
<td>3</td>
</tr>
<tr>
<td>Publication year</td>
<td>1 (early studies)*</td>
<td>.60*</td>
<td>2.1 (Australia)*</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2 (recent studies)*a</td>
<td>.03</td>
<td>2.2 (New Zealand)*</td>
<td>8</td>
</tr>
</tbody>
</table>

Note. SI/NT = sensory integration with no treatment. Sublevel effect sizes and number of contributing studies are shown in parentheses.
*Level and sublevel that attained homogeneity. Level and sublevel that showed significant between variations.
*Significantly different from zero at alpha = .05.

Results of the Sensory Integration Versus Alternative Treatment Analysis

Differences among SI/ALT studies using Qr were non-significant, \( \chi^2(1, 16) = 20.7, p > .10 \), implying that the effect sizes in this comparison were homogeneous. The average effect size was .09 and not significantly different from zero (95% confidence interval = -.11-.28). Thus, the estimated average effect size is representative of the total group of the studies, and no further inferential analysis was performed. The average effect sizes of the measurement categories of SI/ALT studies are presented in Table 5. The magnitude of the effect sizes varied from -.19 (95% confidence interval = -.83-.64) in the language category to .28 (95% confidence interval = -.05-.62) in the motor category. This result implies that sensory integration and alternative treatment approaches showed equal effects.

Chronological Trend of the SI/ALT Studies

Although not significant, a trend appeared in which recent studies yielded smaller effect sizes than the earlier studies in

Table 5
Average Effect Sizes for Measurement Categories by Early and Recent Studies

<table>
<thead>
<tr>
<th>Category</th>
<th>Effect Size</th>
<th>SE</th>
<th>k</th>
<th>N</th>
<th>Effect Size</th>
<th>SE</th>
<th>k</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychoeducational</td>
<td>.62*</td>
<td>.19</td>
<td>11</td>
<td>4</td>
<td>.07</td>
<td>.23</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>Behavior</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>-.08</td>
<td>.33</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Language</td>
<td>.25</td>
<td>.54</td>
<td>3</td>
<td>1</td>
<td>.04</td>
<td>.44</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Motor</td>
<td>.65*</td>
<td>.21</td>
<td>7</td>
<td>4</td>
<td>-.04</td>
<td>.27</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Sensory-perceptual</td>
<td>.41</td>
<td>.64</td>
<td>4</td>
<td>1</td>
<td>.11</td>
<td>.22</td>
<td>28</td>
<td>7</td>
</tr>
</tbody>
</table>

Note. NA = no effect size in this category.
*Significantly different from zero at alpha = .05.
the SI/ALT comparison. Pearson product–moment correlation between publication year and effect size was significant and negative, \( r(16) = -0.73, p = 0.001 \). The related characteristics of the chronological trend were similar to those in the SI/NT comparison. The publication year correlated significantly with the number of outcome measures, \( r(16) = 0.69, p = 0.004 \), and the number of measurement categories, \( r(16) = 0.56, p = 0.033 \).

**Summary of Findings**

**Are Sensory Integration Treatment Approaches Effective?**

Sensory integration treatment approaches were found to be most effective in the earlier studies that compared the performances of sensory integration groups with no-treatment groups, whereas the recent SI/NT studies found no sensory integration effect. Sensory integration effect was statistically equal to the effect of other alternative treatments. For both comparisons, there was a chronological trend of recent decrease of the sensory integration effect.

**What Do Sensory Integration Treatment Approaches Improve?**

In the earlier studies that compared sensory integration groups with no-treatment groups, the improvement occurred in the psychoeducational and motor performance areas. Sensory integration effect was statistically equal to the alternative treatment effects in all areas. No improvement was noted in the sensory-perceptual area. A similar pattern was obtained for the alternative treatment.

**For Whom Are Sensory Integration Treatment Approaches Effective, and What Moderator Variables Should Be Considered?**

The sensory integration effect was statistically indistinguishable for adults with chronic conditions and children with learning disabilities and other developmental disorders. There was no association between the magnitude of sensory integration effect and sensory integration quality of the treatment, meaning that a pure sensory integration treatment method did not show advantages over adapted methods. Likewise, sensory integration did not show notable advantage over alternative treatment methods. With respect to treatment frequency and duration, there was no indication that a better result was achieved by increasing the treatment hours. Two earlier studies with the youngest group of children (about 4 years of age) yielded the largest effect size of .92, but this result could not be attributed to the age factor because of the confounding effect of the chronological trend.

**What Other Factors Influenced the Magnitude of the Effect Sizes of Sensory Integration Treatment Approaches?**

The chronological trend was demonstrated in both SI/NT and SI/ALT comparisons. Comparing the studies from different geographic locations, the studies (all published before 1982) from Australia yielded the largest effect size of .93. However, because of the confounding of the chronological trend, the reason for the large effect cannot be isolated.

**How Are the Results of This Study Related to the Literature Review?**

The results of the current study shed some light on contradictory findings reported by previous reviews. Ottenbacher (1982) reported a positive effect size (.68) of sensory integration treatment approaches as a result of the meta-analysis of the SI/NT studies published between 1972 and 1981, but the literature review of more recent studies found no overall sensory integration effect (Hoehn & Baumeister, 1994; Polatajko et al., 1992). In regard to the sensory integration effect on the outcome categories, the current findings were similar to those reported by the previous literature reviews. Ottenbacher noted the largest sensory integration effect in motor and psychoeducational categories, similar to the current findings with early SI/NT comparison. The authors of the four more recent literature reviews did not find positive effects in any areas.

**Study Limitations**

Because of the limited number of studies reviewed and sparse representation of pertinent information, it was not possible to investigate the difference of sensory integration effect among different types of disabilities, among preschool-age and school-age children, among different school grades, and among different types of alternative treatments. The original studies lacked comprehensive information about age-specific score distributions, specific types of disability, relationships among the multiple outcome measures, provider qualifications, and detailed descriptions of the treatment strategies. When such information was not reported, it was coded as missing.

Another limitation is the lack of interrater verification of the coding process. Other researchers may use different strategies for coding and defining variables. Certain variables required some degree of judgment, especially sensory

---

### Table 6

<table>
<thead>
<tr>
<th>Measurement Category</th>
<th>Effect Size</th>
<th>SE</th>
<th>k</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychoeducational</td>
<td>.12</td>
<td>.23</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>Behavior</td>
<td>.07</td>
<td>.31</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Language</td>
<td>-.19</td>
<td>.33</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Motor</td>
<td>.28</td>
<td>.17</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>Sensory-perceptual</td>
<td>-.11</td>
<td>.21</td>
<td>28</td>
<td>9</td>
</tr>
</tbody>
</table>

*Note. SI/ALT = sensory integration with alternative treatment.*

*All categories have a homogeneous set of effect sizes.*

*Significantly different from zero at alpha = .05.*
integration quality, design quality, and diagnosis and age. For sensory integration quality, nine individual judgments were obtained and summed into an overall score. Although interrater reliability was not obtained, coefficient alpha for this scale was found to be .80. Likewise, five individual judgments were required for design quality, and a coefficient alpha of .75 was obtained. Finally, diagnosis and age required grouping because of the low frequency of a particular diagnosis. In this case, chronic conditions, such as mental retardation and psychiatric disorders, were grouped for adults, and learning disabilities, aphasia, and other developmental disorders were grouped for children. Possibly, other researchers would choose another strategy. However, other variables, such as number of treatment hours, number of outcomes, professional affiliation, and geographic location, required minimal inference.

Conclusion
The two main findings of this analysis that concern occupational therapy practice are (a) the absence of sensory integration effects in recent studies and (b) the equivalence of sensory integration and alternative treatments. The reason for the chronological trend might be associated with some unidentified variations of treatment implementation or with selection of experimental and control participants. One speculation is that the measurement of treatment quality may not have adequately assessed therapists' competence and ability to understand the ongoing psychological and emotional status of the child. Ayres (1973) wrote “such competence represents more than technical proficiency; it approaches an art” (p. 256). It is possible, although unlikely, that in providing objectivity and rigor for examining sensory integration treatment techniques, recent studies discouraged individual variation required for fostering optimal levels of arousal and motivation of each child. Thus, we believe that it is too early to rule out moderator variables, such as therapist's expectancy level, frequency and types of positive and negative feedback to children, and other motivating factors incorporated in the treatment.

The lack of consistent empirical support for the efficacy of sensory integration treatment approaches poses a great concern among occupational therapists. It must be noted, however, that these studies examined the originally delineated sensory integration techniques in isolation and do not accurately represent sensory integration treatment approaches delivered in clinics, where various new treatment techniques are incorporated (Hanschu, 1995; Oetter, 1997; Trotter, Laurel, & Windeck, 1993; Wilbarger & Wilbarger, 1995; Yack, 1989). The finding that various alternative treatments were observed to have positive effects suggests the need to investigate more rigorously the effect of these alternative activities as part of sensory integration treatment strategies. ▲

Acknowledgments
This study was completed as part of the requirements for the first author’s Doctorate Degree in Educational Statistics and Measurement. Rutgers, The State University of New Jersey. In the spirit of meta-analysis, the coding form and other technical data can be obtained from the first author to evaluate the procedures of this study.

References


