Research on the Efficacy of Sensory Integration Therapy: Past, Present and Future

Han M. Leong and Mark Carter*

Macquarie University Special Education Centre, Macquarie University, Sydney, Australia

Research on the efficacy of sensory integration therapy (SIT) is addressed in this article. Initially, past key reviews of intervention studies until 1994 are considered. Subsequently, more recent studies from 1994 until 2007 are examined. Consistent with numerous previous reviews, no robust evidence supporting the efficacy of SIT was found. Alternative and more parsimonious explanations for purported effects of SIT are considered. In light of the accumulated lack of evidence for the effectiveness of SIT, continued use of the technique outside of research contexts does not appear to be justified.

Sensory integration therapy (SIT) was initially researched by Ayres in the 1960s, and remains a controversial yet highly enduring practice (Hyatt, Stephenson, & Carter, in press; Smith, Mruzek, & Mozingo, 2005). SIT has been advocated for a remarkably wide variety of disabilities and problems. This includes students with learning disabilities, intellectual disability, autism spectrum disorders (ASD) (Smith, Press, Koenig, & Kinnealey, 2005), infants born at risk and/or with regulatory disorders, children from environmentally deprived backgrounds (Schaaf & Miller, 2005), patients with dementia (Robichaud, Hebert, & Desrosiers, 1994), children with developmental coordination disorder (Davidson & Williams, 2000), severely traumatised children and adolescents (Alers, 2005) and adopted children who had been institutionalised for long periods (Lin, Cermak, Coster, & Miller, 2005). While the major focus in the early decades of research on SIT was in the learning disabilities area, more recent studies have increasingly addressed individuals with ASD.

SIT is aimed at improving functional outcomes by correcting proposed sensory integrative dysfunction and addressing ‘underlying neurological processing’ of sensory information (Smith, Press et al., 2005). It is theorised that improving neurological sensory processing will lead to gains in functional outcomes because sensorimotor development is seen as an important substrate for learning (Schaaf &
The direct teaching of specific skills is not a goal of SIT nor is it a part of therapy. Ottenbacher (1982) noted that:

in order to be classified as sensory integration therapy the treatment could not involve desk activities, speech training, reading lessons, or specific perceptual motor skills training. The stated or implied goal of therapy was to improve the way the brain processed and organised sensation, not to teach specific academic or motor skills. (p. 572)

Hence, SIT is not meant to replace formal classroom instruction, but rather, is regarded to be a supplementary treatment that enhances the students’ ability to learn (Schaaf & Miller, 2005). Thus, SIT is most commonly advocated as a significant precursor to learning higher order skills (Ottenbacher, 1982; Schaaf & Miller, 2005; Vargas & Camilli, 1999).

‘Classic’ SIT involves delivery of intervention in a clinical session (Baranek, 2002). SIT is usually implemented by occupational therapists and involves providing controlled sensory stimuli, typically vestibular, tactile and proprioceptive stimulation. The vestibular sense relates to movement (such as spinning) affecting receptors in the middle ear, the tactile sense involves touch, and proprioception relates to the position of muscles and joints. Stimulation is provided to the individuals in order to produce an adaptive response, where they modify their behaviour appropriately in response to therapy (Schaaf & Miller, 2005). These adaptive responses are proposed to improve sensory processing and thus lead to other outcomes (Schaaf & Miller, 2005). Characteristically, therapists will provide various challenges for the student within a play environment and continuously adjust the challenge according to the student’s response to stimulation (Schaaf & Miller, 2005). A wide variety of activities may be involved in therapy, such as the use of hammocks and scooter boards to provide stimulation, rubbing and brushing of the body with various textures, use of weighted vests, and manual compression of joints (see Ayres, 1972; Hoehn & Baumeister, 1994; Smith, Mruzek et al., 2005). ‘Sensory diets’ are a related and more recent intervention involving activities and environmental adjustments, designed to complement the individual’s sensory needs (Smith, Mruzek et al., 2005).

According to a recent survey by Green, Pituch, Itchon, Choi, O’Reilly, and Sigafoos (2006), 38.2% of parents of children with autism currently use sensory integration therapy (ranking as the third most widely used intervention) and 33.2% of parents had used it previously (ranking first). Further, Case-Smith and Miller (1999) reported that SIT was the most commonly implemented intervention used by occupational therapists with children with autism, with 95% of therapists using the intervention at least sometimes. Noting the apparent wide use of SIT, it is important to recognise that it can be resource intensive. A well-trained occupational therapist is needed to observe and assess the individual’s sensory needs, and this therapist will need time to both assess and implement therapy (Baranek, 2002; Schaaf & Miller, 2005). Vargas and Camilli (1999) reported that the therapy they reviewed averaged 60 hours, ranging up to 180 hours. Thus, the resources in money, time and clinical facilities consumed in order to deliver SIT can be significant.
In view of the continued wide application of SIT and its ongoing controversial status, a summary of reviews of older studies is warranted, as is an examination of more current research. In this article, key reviews of research on the efficacy of SIT are summarised and discussed, starting with Ottenbacher’s (1982) meta-analysis, up until the meta-analysis of Vargas and Camilli (1999), covering studies from 1972 to 1994. The Vargas and Camilli (1999) study has been chosen as a point of reference as it stands to date as the largest major published meta-analysis on the efficacy of SIT. More recent studies, published after 1994 and until 2006, are then reviewed in order to examine whether there have been significant advances in research into SIT. Finally, the results of the review and implications are discussed.

**Research Prior to 1994**

Ottenbacher (1982) reviewed eight studies published between 1972 and 1981. The total sample size was 317 participants, out of which 89 had a diagnosis of intellectual disability, 191 were diagnosed as learning disabled, 18 were diagnosed as aphasic, and 19 were diagnosed as ‘at risk’ for reading disorders. From these eight studies, Ottenbacher (1982) identified 47 outcome measures that yielded an average effect size of 0.79. In educational terms, this is a large effect size. Further analysis of the results yielded that the mean effect size was highest for motor-reflex variables (1.03) and lowest for language (0.52). Mean effect size was highest for participants diagnosed as ‘at risk’ or aphasic (1.2) and lowest for participants with intellectual disability (0.52). Based on these early results, Ottenbacher (1982) concluded that the application of SIT had significant effects on learning outcomes. While these early findings were encouraging, the Ottenbacher (1982) review has been criticised by Hoehn and Baumeister (1994) and Arendt, MacLean, and Baumeister (1988) on a number of grounds. These included the design issues in the primary studies themselves, such as lack of comparison of experimental treatment groups to alternative treatment control groups, as well as lack of control for bias arising from use of multiple outcomes from the same studies.

Whereas Ottenbacher (1982) examined studies of SIT across diagnostic categories, a number of reviews have been conducted to evaluate the effect of SIT on specific groups. Arendt et al. (1988) reviewed eight studies of SIT used on students with intellectual disabilities published between 1977 and 1983, and found significant methodological flaws in the research, including three studies that used anecdotal and descriptive measures for reporting results. While all but one study reported positive results in favour of SIT, there were serious experimental design flaws in the studies. The observed effects could be attributed to a range of factors including Hawthorne or placebo effects, maturation, observer bias and possible positive reinforcement of other behaviour. Arendt et al. (1988) concluded that ‘until the therapeutic effectiveness of sensory integration therapy with mentally retarded persons is demonstrated, there exists no convincing empirical or theoretical support for the continued use of this therapy with that population outside of a research context’ (p. 409).
Schaffer (1984) reviewed five key outcome studies that employed SIT with individuals with learning difficulties, including three studies by Ayres that were also included in Ottenbacher’s meta-analysis. Schaffer (1984) examined the design of the studies and found them ‘fraught with serious methodological errors’ (p. 76) due to lack of random assignment, absence of comparison to alternative intervention groups, lack of blind testing, non-equivalent groups, lack of clear reliability data, and unclear definitions of participant populations. Schaffer (1984) advised that SIT’s efficacy could not be judged based on these key studies due to the fundamental problems related to internal validity.

Hoehn and Baumeister (1994) reviewed seven studies on SIT, also with students with learning disabilities, which were published between 1984 and 1992. The outcomes examined fell into four groups: postrotary nystagmus; sensorimotor, perceptual, and motor performance; cognitive, language and academic measures; and self-esteem/self-concept, attention and behavioural measures. Postrotary nystagmus refers to lateral movement of the eyes present after a person has been spun and this measure is important as it has been used by researchers as an index of the level of proposed sensory integrative dysfunction. For postrotary nystagmus outcomes, only one out of five different experiments had significant consistent gains. For sensorimotor, perceptual, and motor measures, no consistent significant benefits were reported although two out of six studies documented mixed results. For the last two outcome groups, no significant unique improvements associated with SIT were reported. Hoehn and Baumeister (1994) concluded that no evidence was found for the efficacy of SIT for groups with learning disabilities.

The Vargas and Camilli (1999) study is the most comprehensive review of SIT to date in a peer refereed journal. A meta-analysis of 26 studies, published from 1972 to 1994, was conducted, including many studies in previously discussed reviews. The studies providing a comparison of experimental to no treatment control groups involved a total of 341 experimental participants and 237 control participants. Where the experimental group was compared to alternative treatment groups, total sample sizes were 250 participants in experimental groups and 191 in comparison groups. Effect sizes were weighted to reduce bias when multiple measures were taken from the same study. Studies comparing experimental groups to no treatment groups had a weighted average effect size of 0.29, which although statistically significant, is relatively low for educational purposes. The common threshold for a clinically significant effect size for educational purposes is around a quarter to a third of a standard deviation (see Forness, 2001; What Works Clearinghouse, n.d.). Vargas and Camilli (1999) examined the studies further and reported that studies published earlier (1972–1982) had a significant mean effect size of 0.6, but newer studies (published from 1982 to 1994) had a non-significant mean effect size of only 0.06. In studies that compared experimental groups to alternative treatment groups no significant difference was found in any area. Similarly, the quality of SIT treatment and total treatment hours could not be statistically associated with better results, providing no evidence of a dosage effect.
In summary, while the findings of Ottenbacher (1982) suggested that SIT was an effective intervention, subsequent reviews disconfirm these findings. Later analyses by Arendt et al. (1988), Schaffer (1984) and Hoehn and Baumeister (1994) provided no compelling evidence for the efficacy of SIT based on the studies reviewed. Vargas and Camilli (1999) confirmed the findings of Ottenbacher (1982) with regard to older studies but the results were not maintained in later research. Vargas and Camilli (1999) found limited effects in studies overall and no evidence that SIT was effective compared to alternative treatments. Thus, reviews up to the Vargas and Camilli (1999) meta-analysis suggest that SIT is unproven to be effective in improving learning outcomes. Nevertheless, as previously noted, SIT has continued to be advocated and used. The question arises as to whether more recent research on sensory integration has in fact provided more substantive evidence that the intervention is effective.

Research Post-1994

Studies from 1994 until 2007 will now be reviewed in order to determine if there has been substantive change in the status of research on sensory integrative therapy in more recent years.

Search Method

A search was conducted in March 2007 using the descriptor ‘Sensory Integrat$’ ('$' representing truncation) to search the PsycInfo, A+, and ERIC databases for articles published later than 1994. In addition, reference lists of articles were used to identify further relevant research. Only articles from peer-refereed journals were included. The criteria used for selecting studies were that the study provided empirical data on outcomes, the authors identified the intervention as involving SIT and the intervention was consistent with the theoretical framework discussed earlier in this article (i.e., focusing on altering fundamental neurological processing rather than skill development). Consistent with previous reviews, studies were selected that involved direct delivery of SIT in a clinical treatment session, an approach described as ‘classic’ sensory integration by Baranek (2002). Excluded were isolated interventions, such as use of weighted vests and pressure garments, that were aligned with a sensory integrative approach but did not involve direct delivery of SIT in a clinical treatment session. Studies reviewed by Vargas and Camilli (1999) were also excluded. A total of eight studies was selected using these criteria. Summaries of the studies are provided in Table 1.

Review of Studies

One study examined the response of individuals with intellectual disabilities to SIT. Soper and Thorley (1996) conducted an investigation involving adults who had
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Design</th>
<th>Treatment length/ frequency</th>
<th>Outcome measures</th>
<th>Claimed results</th>
<th>(1) Interrater reliability and (2) Procedural reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Smith &amp; Bryan (1999)</td>
<td>5 boys with autism, aged 4–5 years.</td>
<td>AB</td>
<td>30 min. interventions sessions (frequency not stated) for 10-week period plus consultation with pre-school staff.</td>
<td>10 min weekly play session. Mastery and non-mastery play, non-engaged behaviours, interaction with peers and adults.</td>
<td>Improvements in mastery play and non-engaged behaviours. Results for adult interaction were variable. No change in interaction with peers. Non-mastery play data not presented.</td>
<td>(1) 90% reliability on 29% of observations. Observers blind to condition. (2) Not reported.</td>
</tr>
<tr>
<td>Davidson &amp; Williams (2000)</td>
<td>37 children with developmental coordination disorder.</td>
<td>Pre-test–post-test single group with retrospective analysis of data.</td>
<td>10-week block of therapy (frequency and duration of sessions not stated) followed by 1-year school and home programs.</td>
<td>Movement ABC test (manual dexterity, ball skills and balance) and Beery-Buktenica Developmental Test of Visual-Motor Integration.</td>
<td>Statistically significant but limited improvements only in fine motor skills and visual-motor integration. Therapy was deemed relatively ineffective.</td>
<td>(1) Not reported but scales have known psychometric properties. Raters not blind. (2) Not reported.</td>
</tr>
<tr>
<td>Green et al. (2003)</td>
<td>(a) 34-year-old with intellectual disability and possible ASD. (b) 28-year-old with severe intellectual disability and autism. Both had maladaptive behaviours suspected to arise from sensory processing deficits.</td>
<td>(a) ABAB (b) ABA</td>
<td>(a) 4-week treatment phases, twice weekly × 50 min. (b) 4-week treatment phase, thrice weekly × 30 min.</td>
<td>(a) Screaming, participation and tolerance of physical contact in a contrived activity. (b) Covering ears, deep breathing and tapping during same time of day for pre-established routine.</td>
<td>(a) 75% reduction in incidents of screaming by the end of the second treatment phase, no significant improvement in participation and some evidence of improvement in tolerance of physical contact and interaction with environment. (b) No significant improvement in any behaviours.</td>
<td>(1) 85% agreement, raters blind to experimental periods but no detail on proportion of sessions. (2) Not reported.</td>
</tr>
<tr>
<td>Linderman &amp; Stewart (1999)</td>
<td>2 boys, 3-years-old with pervasive developmental disorders.</td>
<td>AB</td>
<td>(a) 1 hr per week for 11 weeks. (b) 1 hr per week for 7 weeks.</td>
<td>(a) Social interaction, approach to new activities and response to holding and hugging for first child. (b) Social interaction, functional communication and response to movement</td>
<td>(a) Positive results for all outcomes. (b) Significant positive results in social interaction and response to movement but not functional communication.</td>
<td>(1) Mean reliability 66.5% (33–100%) for participant 1 and not reported for participant 2. (2) Not reported.</td>
</tr>
</tbody>
</table>
Study | Participants | Design | Treatment length/ frequency | Outcome measures | Claimed results | (1) Interrater reliability and (2) Procedural reliability |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith, Press et al. (2005)</td>
<td>7 participants 8–19 years with pervasive developmental disorder/ intellectual disability.</td>
<td>Group design with alternating of treatment conditions (SIT and table-top activities).</td>
<td>2 × 1 week treatment phases per condition within a 4-week study period, 30 min. sessions.</td>
<td>Frequency of self-stimulatory/self-injurious behaviour within school routine, interval recording.</td>
<td>Average of 11% decrease in treatment phases for SIT.</td>
<td>(1) Correlation of 0.92 for between 3 raters and original scorer for 10% of segments, raters blind to phases of treatment. (2) Not reported.</td>
</tr>
<tr>
<td>Soper &amp; Thorley (1996)</td>
<td>28 participants with intellectual disability, limited language and tactile defensiveness.</td>
<td>Group design comparing SIT to sensory stimulation.</td>
<td>Once weekly for 1 hour for 9 months. Staff to client ratio of 1:1 or 2:1.</td>
<td>(a) Clinical Observations Checklist addressing sensory tolerance, adaptation and postural responses. (b) Behavioral Checklist with items influenced by the Sensory Integration Inventory for Adults. (c) Ayres Scale of Adaptive Responses.</td>
<td>Experimental group improved more than controls across a range of measures.</td>
<td>(1) Inter-rater reliability of 92% for Clinical Observations Checklist on 4 participants. No reliability reported on other measures. (2) Not reported.</td>
</tr>
<tr>
<td>Urwin &amp; Ballinger (2005)</td>
<td>5 adults aged 19–65 years with moderate to severe intellectual disabilities and inappropriate responses to tactile stimulation. 3 had ASD, 1 had autistic traits and 2 had epilepsy.</td>
<td>ABA</td>
<td>1 × 4 week treatment phase, 0-40 min. sessions.</td>
<td>Within context of horticulture tasks: Level of engagement, level of maladaptive behaviour and functional behaviour goal level on Goal Attainment Scale.</td>
<td>Evidence for improvements in all areas, particularly maladaptive behaviours.</td>
<td>(1) Calculated on 10% of session using blind raters. Kappa ranged from .71 to .91. (2) Not reported.</td>
</tr>
<tr>
<td>Robichaud (1994)</td>
<td>40 patients with dementia aged 60+ years.</td>
<td>Control group with randomised assignment.</td>
<td>3 times weekly (30–45 min.) over 10-week period.</td>
<td>Frequency of disruptive behaviour and reactions by caregivers (Revised Memory and Behavior Problems checklist, French version), level of assistance needed for activities for daily living (Psychogeriatric Scale of Basic Activities of Daily Living).</td>
<td>No significant difference between study and control group</td>
<td>(1) Not reported but scales have no psychometric properties. Only first collection of data was blind. (2) Not reported.</td>
</tr>
</tbody>
</table>
limited language skills and ‘apparent tactile defensiveness’. The measures used in this study warrant comment as they were rating scales related to sensory integration, and direct measures of functional outcomes were not undertaken. While significant improvements were reported on most SIT-based measures, with greater gains associated with the SIT group, these data offer little insight into whether this had real impact on functional outcomes for participants. Data on changes in self-stimulatory and self-injurious behaviours were embedded in the tactile, vestibular and proprioceptive outcomes of the behavioural checklist, but gains appeared to be relatively small and not significant for tactile outcomes. Further, the clinical significance of changes was not interpretable as direct measures of behaviour (e.g., frequency or duration) were not reported. It was not explicitly stated that assignment to groups after matching was random or that assessors were blind to group assignment. Further, inter-rater reliability was only conducted with 4 of 28 participants for one of the three checklists and procedural reliability was not assessed. Thus, this study appears to have significant methodological problems and provides no clear demonstration that the intervention improved any functional outcome for participants.

Sensory processing difficulties are often described as a key feature of ASD (O’Neill & Jones, 1997). Five studies (Case-Smith & Bryan, 1999; Green et al., 2003; Linderman & Stewart, 1999; Smith, Press et al., 2005; Urwin & Ballinger, 2005) provided examination of the effects of SIT on individuals with ASD or combined intellectual disability and ASD. Linderman and Stewart (1999) implemented SIT with two 3-year-old boys who were diagnosed with autism or were reported to show symptoms of autism. Study outcomes were measured using the Functional Behavior Assessment for Children with Sensory Dysfunction and this was modified, with the scale expanded from 4 to 10 points. Again, a rating scale was used to evaluate outcomes rather than a direct measure of functional behaviour and, particularly with the (poorly documented) modifications, these data are virtually impossible to interpret in terms of meaningful outcomes. For the first participant, results were confounded by the fact that a special education pre-school program had been started early in baseline and speech therapy in the third week of intervention. Results were interpreted by Linderman and Stewart (1999) to be probably clinically significant as the participant had already made gains in the intervention phase, before the speech therapy was introduced. Results for the second participant showed apparently limited gains on two of the three dependent variables, but again it should be stressed that the functional implications of these results are opaque.

Due to the autocorrelation in their data, which is common in repeated measures research, Linderman and Stewart (1999) attempted to analyse most of their data using the two standard deviation bands method (where change of two standard deviations above baseline was proposed to suggest significant improvements). This was problematic, however, as the standard deviations were very small because they were calculated from the short and very stable baseline stages. Standard deviation-based effect sizes are often used for evaluating the strength of treatment in group
designs but rarely in single-subject research. It would be reasonable to expect much less variation in repeated measures of the behaviour of an individual, than between individuals. Consequently, interpretation of the two standard deviation criterion is questionable. Large effect sizes can simply be an artefact of very stable baselines and very small standard deviations, as was the case here. Weak AB designs were employed for both participants and inter-rater reliability was poor. The mean exact agreement for the first participant was 66.5% (range 33–100%) and no inter-rater reliability was taken for the second participant. Procedural reliability was not reported. Most importantly, and as previously noted, the use of an adapted rating scale makes the results difficult to interpret in terms of functional outcomes, particularly as the meaning of points on the scale was not explained.

Smith, Press et al. (2005) conducted a study with children and adolescents described as having pervasive developmental disorders and/or severe intellectual disability, although participant description was scant. The frequency of self-stimulatory or self-injurious behaviour was measured in classrooms prior to treatment each day, immediately post SIT or table-top activity sessions, and one hour after these sessions. In addition, teachers were asked to provide a subjective rating of levels of relevant behaviour each day on an ordinal scale. While only conducted on 10% of sessions, the level of inter-observer reliability was sound and raters used for reliability were reported as blind to the treatment condition. Smith, Press et al. reported that self-stimulatory behaviours during the school routine decreased by an average of about 11% after a 3-minute SIT session and increased about 2% following table-top activities. Both represent relatively small changes considering the type of behaviour measured. While the correlation between frequency of SIT and teacher impressions of behaviour was statistically significant, it was modest (r = 0.32), providing only limited support to the findings. The timing of the measurement of outcomes soon after treatment also warrants comment. It seems entirely plausible that transient effects, say associated with antecedent physical exercise or sensory stimulation, may influence subsequent behaviour when measured relatively soon after intervention. Further, the SIT and table-top treatments were conducted in alternate weeks so, presumably, the effects of SIT completely dissipated from one week to the next. Thus, only relatively limited and very short-term effects were demonstrated, the study was of short duration and results were inconsistent across participants (frequency of challenging behaviour did not decrease for some participants). Consequently, the results of this study provide very little evidence of the efficacy of SIT in reducing the frequency of stereotypic and self-injurious behaviour.

Urwin and Ballinger (2005) conducted a study with adults who had intellectual disability and were also diagnosed with ASD in most cases. The effects of 10–40 minute individualised SIT were examined on engagement, maladaptive behaviour and Goal Attainment Scale scores, in an immediately following 10-minute horticultural task. No operational definitions were offered of key dependent variables and while duration measures were apparently employed for engagement
and maladaptive behaviour, insufficient detail was provided of the procedures to allow replication. Despite claims of statistically significant differences across phases, inspection of presented graphs reveals poor and inconsistent demonstration of experimental control, particularly in regard to engagement and maladaptive behaviour. Interpretation is complicated by high variability, modest and inconsistent effects, and failure of dependent variables to return to baseline levels during the withdrawal phases. Again, only very immediate effects of the intervention were addressed. Further, even if the effects were credible, the cost–benefit of investing up to 40 minutes in therapy to gain improvement in 10 minutes of subsequent activity must be open to question.

Green et al. (2003) also conducted a study involving two adults with intellectual disability who had autism or suspected ASD. Any attempt to interpret this research should be regarded with caution as the description of the procedures for data collection was unclear and insufficient to allow replication. For participant 1, data were confusingly presented on the ABAB design. There appeared, however, to be no decrease in scream frequency in the first treatment phase, with a 50% decrease in the subsequent withdrawal phase and a further 25% decrease in the final treatment phase. The authors claimed a 75% decrease in screaming in the study but appear to have failed to recognise that these changes only started after the initial treatment was withdrawn, not when it was introduced. That is, effects were more strongly associated with the withdrawal of intervention than its introduction. While no data were presented, the authors acknowledge that there was no change in ratings of withdrawal/engagement behaviour. Claimed improvements in Goal Attainment Scale measures were not interpretable as data were only reported in treatment phases. Despite the (inappropriate) application of multivariate parametric procedures to the analysis of these data and anecdotal reports of positive behaviour change, there was absolutely no convincing evidence of any consistent effect of the intervention. In relation to the second participant, the authors acknowledge that the intervention had no effect, a conclusion consistent with the data presented.

Case-Smith and Bryan (1999) examined the use of SIT for five boys with autism. Unfortunately, frequency of the 30-minute therapy sessions was not stated and temporal proximity of therapy to the 10-minute period of pre-school free play, used to assess behaviour, was also unclear. These factors complicate interpretation. There appeared to be intervention-related increases in mastery play for four participants, although intervention effects were delayed in two cases. Interpretation of decreases in non-engaged play were complicated by sharply decreasing baselines in two cases and intervention effects were mildly delayed in the remaining instances. Changes in adult intervention were variable and there was little evidence of change in child interaction. The authors did not present data on non-mastery play, claiming it was difficult to interpret. There was certainly modest evidence of changes in performance concomitant with introduction of SIT for some children but this was variable across children and dependent variables. Given the weak internal validity of the AB designs employed and the limited data collected, these results should be treated with
caution. While certainly not a strong research design, unlike several other studies reviewed, Case-Smith and Bryan (1999) did clearly define functionally relevant dependent variables and conducted reasonable inter-observer reliability checks.

Davidson and Williams (2000) examined a combined SIT and perceptual-motor training therapy for children with developmental coordination disorder. A retrospective analysis of data was conducted prior to the program and after a year. Statistically significant improvements were only reported for fine motor skills and visual motor integration, and even these improvements were very small. The weak nature of the study design was such that even the small changes observed over a year could not reasonably be attributed to treatment effects. Davidson and Williams (2000) recognised the methodological weaknesses inherent in the study but concluded that the therapy was likely to be relatively ineffective.

In an unusual study, Robichaud et al. (1994) examined the efficacy of SIT on senior patients with dementia. A randomised control group design was used but assessors were only blinded to assignment at pre-test. At post-test, both the study group and control group improved and no significant difference between the groups was found. Among the reasons given for explaining the lack of efficacy of intervention, Robichaud et al. (1994) noted that SIT may not be appropriate for treating dementia, which is an irreversible disorder.

**Discussion**

Many of the studies examined suffered from numerous methodological problems, making interpretation difficult. Measurement scales that could not be interpreted in terms of functional changes in behaviour were employed in some studies (Linderman & Stewart, 1999; Soper & Thorley, 1996), only very immediate effects were investigated in others (Smith, Press et al., 2005; Urwin & Ballinger, 2005) and inter-observer reliability was inadequate in others (Linderman & Stewart, 1999; Soper & Thorley, 1996). The studies employing small n designs were particularly poorly implemented and interpreted for the most part, as previously documented. Problems with interpretation were well illustrated in the study of Green et al. (2003) where a 75% decrease in screaming was reported but the researchers did not appear to recognise that this was more strongly associated with the withdrawal of intervention than its introduction and that experimental control was not demonstrated.

In the sample of studies examined, SIT was applied to a remarkably wide variety of problems ranging from ASD to dementia. Subject description was universally limited, often amounting to little more than a diagnostic label (e.g., Davidson & Williams, 2000; Linderman & Stewart, 1999; Robichaud et al., 1994; Smith, Press et al., 2005; Urwin & Ballinger, 2005). Absence of clear participant description would obviously severely hamper any attempt to determine if the effectiveness of SIT were associated with specific participant characteristics. Further, participants in some studies were selected because they demonstrated a variety of different specific characteristics such as tactile defensiveness or sensory modulation disorder in the
tactile domain (Soper & Thorley, 1996; Urwin & Ballinger, 2005), poor sensory processing (Green et al., 2003), disruptive behaviour (Robichaud et al., 1994) or stereotypic and self-injurious behaviour (Smith, Press et al., 2005), while in other studies, the intervention appeared to be applied to any participant within a diagnostic group (Case-Smith & Bryan, 1999; Davidson & Williams, 2000). Davidson and Williams (2000) note that ‘for the efficacy of any intervention to be demonstrated, both the clinical condition and the client group must be defined’ (p. 496). This advice would be well heeded by those researching in the area.

It was also interesting that none of the studies reported formal checks of procedural reliability, even those employing small $n$ designs, where such checks are considered important indicators of research quality (see Horner et al., 2005). Smith, Press et al. (2005) argued that SIT is difficult to operationally define, possibly accounting for the failure to monitor treatment integrity. If so, then interpretation of this line of research will present an ongoing problem, as it will remain uncertain as to whether the intervention was implemented as intended.

Nevertheless, the present examination of more recent studies of SIT provides no substantive evidence that the intervention improves functional outcomes for any diagnostic group. Where positive results were reported, they were typically modest and inconsistent across participants and dependent variables. These findings are consistent with numerous previous reviews (Arendt et al., 1988; Baranek, 2002; Dawson & Watling, 2000; Hoehn & Baumeister, 1994; National Research Council, 2001; Perry & Condillac, 2003; Roberts, 2004; Shaw, 2002; Vargas & Camilli, 1999) in indicating that there is little evidence to support the efficacy of SIT. Vaughn and Linan-Thompson (2003) observed that while underlying neurological disorders or processing disorders may exist in certain populations with disabilities, researchers and educators have not been successful at reliably assessing these disorders and designing treatments that prove to be effective. This observation appears to hold true for the treatment of purported underlying neurological sensory processing disorders using SIT.

Even if it were accepted that the inconsistent and modest changes reported in some of the studies were of clinical importance, several significant problems of interpretation exist and explanations other than SIT are possible for observed effects. In early studies of SIT, particularly with individuals with learning disabilities, the expected definitive outcome of addressing underlying neurological dysfunction was improvement in higher order skills, such as academic performance. If substantial changes in such performance could be unambiguously attributed to SIT, an argument could certainly be mounted that underlying sensory processing may have been affected. The focus of some of the more recent studies examined here was quite different. In these cases, very immediate short-term effects on behaviours such as task engagement (Urwin & Ballinger, 2005) and stereotyped or self-injurious behaviour (Smith, Press et al., 2005) were examined. It is certainly not unreasonable to expect that the sorts of activities described in a SIT session might have short-term effects on behaviour, minutes or hours after intervention. For example, most of the
therapy used by Urwin and Ballinger (2005) was described as calming and inhibitive, including gentle swinging, vibration and deep pressure, and similar techniques were employed commonly across other studies. It seems entirely plausible that such calming stimulation could affect behaviour in the short term and there seems to be little reason to attribute such purported effects to fundamental correction of underlying neurological sensory processing deficits. If any argument is to be forwarded that changes in neurological processing are implicated in SIT, a starting point would be to demonstrate robust and long-term change in behaviour. Further, Mason and Iwata (1990) provided a clear demonstration that apparent effects of SIT on challenging behaviour can be artefactual, relating to such factors as non-contingent positive reinforcement (e.g., approval, physical contact or provision of desired sensory stimulation) or negative reinforcement (e.g., escape from task demands). Thus, before any purported effects are ascribed to the remediation of underlying neurological deficits, more mundane and parsimonious explanations should be excluded.

A number of researchers attempted to address reported sensory problems in autism and other disabilities. If SIT does affect responses to sensory stimuli, effects may simply reflect behavioural systematic desensitisation to those stimuli, as might be employed in an intervention addressing phobias or anxiety (see Mazur, 1994). For example, Soper and Thorley (1996) stressed the ‘gentle, gradual introduction of what were, for most clients, novel experiences related to touch and movement’ (p. 476), an approach consistent with the interventions described by others (e.g., Linderman & Stewart, 1999; Smith, Press et al., 2005; Urwin & Ballinger, 2005). Any effects observed as a product of such interventions may well be more parsimoniously explained in terms of respondent extinction associated with gradual exposure and desensitisation to stimuli. Similarly, in some cases the interventions might just involve removing or modifying antecedents that trigger behaviours, which is in fact another generic and widely employed behavioural intervention strategy. SIT used for the sole reason of calming (or increasing arousal) of learners prior to taking part in activities is questionable and there may be other less resource-intensive ways to achieve the same short-term effects.

Thus, SIT interventions may in fact be unsystematically adopting strategies that have the potential to be effective, such as desensitisation. These strategies are probably better applied systematically, possibly accounting for the lack of convincing and robust treatment effects for SIT. The historically poor track record of SIT may indicate that the theoretical framework is fundamentally flawed. Even if SIT does result in functional behaviour change, and this has not been established, it is important to accurately identify why the intervention may be effective and select the most appropriate theoretical framework to explain outcomes. The SIT framework does not appear to be parsimonious and quite possibly obfuscates the effects of more systematic and established strategies. Researchers working in the area of SIT need to compare the intervention to other established strategies to determine if it makes any unique contribution.
Recent years have witnessed a shift in SIT research, with a greater emphasis on ASD. Relative to, say, learning disability, ASD is a low-incidence condition characterised by a high degree of heterogeneity. Consequently, small n research designs have often been employed. Based on the studies reviewed here, several recommendations can be offered to future researchers. Researchers need to clearly describe participants and the criteria for selecting them for intervention. In addition, outcome variables must be both interpretable and functional. There appears to be an increasing focus on behavioural outcomes in recent research. Such behaviour can plausibly be influenced by immediate antecedent events (e.g., rocking a child may settle behaviour). As a starting point there is a need to demonstrate that intervention has sustained impact if such behavioural changes are to be attributed to alteration in underlying neurological processing, as suggested in SIT. Further, there are well-validated intervention approaches to many of the problems addressed by SIT (e.g., treatments of self-injurious and stereotypic behaviour based on functional analysis). Researchers need to examine whether SIT has comparable efficacy with such interventions. Finally, stronger research designs need to be employed (e.g., multiple baseline) with due attention to the basic mechanics of small n research. This includes establishing adequate inter-observer reliability and procedural integrity as well as a conservative approach to interpretation of graphic data.

There is no doubt that there has been a move in special education toward evidence-based practice in recent years (e.g., Gersten et al., 2005; Horner et al., 2005; Odom et al., 2004). Adoption of this paradigm means the profession must now develop responses to evidence. The major question regarding SIT is whether there is sufficient evidence to indicate it is disproved or whether it remains unproven. On balance, and considering all the accumulated evidence, a strong case can be put that at this stage SIT appears to be an ineffective intervention in terms of the standards typically adopted in special education. If this is accepted, the response seems clear. Continued clinical application of SIT should not be supported and resources should be diverted to interventions with a stronger evidence base. If the most generous interpretation of the available evidence is adopted, it might be concluded that SIT is an unproven intervention. If this position were accepted, two options for response would seem to be available. First, it could be argued that SST is apparently a widely used and established intervention and clinical application should continue, perhaps cautiously, until it is proven ineffective. Such an argument reverses the normal burden of scientific proof (Hyatt et al., in press) and paves the way for continued use of other existing but unproven practices and, by extension, the adoption of new ones until such time as they are disproved. It is difficult to see how this could be construed as consistent with an evidence-based approach or how it could improve educational practice. The second option is to discontinue clinical use of SIT until such time as it can be demonstrated to be effective. In recommending this option, it should be noted that there are empirically supported alternative interventions for the types of functional problems SIT
purportedly addresses. Thus, there appears to be little risk of a treatment vacuum being created. The response of the educational community to dilemmas of this type over the coming years will provide an interesting test of commitment to evidence-based principles.

The ethical issue of withholding treatment was raised in a number of the studies reviewed (Case-Smith & Bryan, 1999; Davidson & Williams, 2000) and in one instance (Case-Smith & Bryan, 1999) specifically in regard to justification for the use of a weak research design. There is, however, a more important ethical question. Is it ethical to continue providing what is at best an expensive, controversial and unproven therapy, and at worst a demonstrably ineffective intervention, to individuals with disabilities? At the very least, practitioners of SIT should always inform parents and other decision makers that there is insufficient research evidence to demonstrate that SIT is effective. If the decision is made to use SIT despite the current state of evidence, professionals should strongly consider monitoring the effects of therapy systematically and continually assess the outcomes of treatment (Baranek, 2002; Smith, Mruzek et al., 2005). Care should also be taken to make sure that SIT does not take away time from proven interventions (Smith, Mruzek et al., 2005) and practitioners should take steps to minimise clinical risks that may arise due to unintended artefactual effects (Mason & Iwata, 1990; Smith, Mruzek et al., 2005).

In conclusion, in light of the accumulated lack of evidence for the effectiveness of SIT, continued use outside the context of research appears to be unjustified. SIT therapy is currently being applied for very different populations, including learning disabilities, intellectual disabilities and ASD. SIT has yet to be proven to be effective with any of these groups even though research dates back around 40 years. Until research can sufficiently indicate otherwise, practitioners are advised to seek proven alternative treatments for clients with disabilities.

References


Copyright of Australasian Journal of Special Education is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder’s express written permission. However, users may print, download, or email articles for individual use.