



The Open Journal of Occupational Therapy

Volume 9
Issue 1 *Winter 2021*

Article 8

January 2021

Inter-Rater Reliability of Goal Attainment Scaling with Children with Sensory Processing Disorder

Teresa A. May-Benson
SPIRAL Foundation – USA, tmay-benson@comcast.net

Sarah A. Schoen
Rocky Mountain University of Health Professions – USA, sarah.schoen@spdstar.org

Alison Teasdale
SPIRAL Foundation – USA, research@thespiralfoundation.org

Jane Koomar
OTA the Koomar Center – USA, jane.koomar@gmail.com

Follow this and additional works at: <https://scholarworks.wmich.edu/ojot>

 Part of the Occupational Therapy Commons

Recommended Citation

May-Benson, T. A., Schoen, S. A., Teasdale, A., & Koomar, J. (2021). Inter-Rater Reliability of Goal Attainment Scaling with Children with Sensory Processing Disorder. *The Open Journal of Occupational Therapy*, 9(1), 1-13. <https://doi.org/10.15453/2168-6408.1693>

This document has been accepted for inclusion in The Open Journal of Occupational Therapy by the editors. Free, open access is provided by ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.

Inter-Rater Reliability of Goal Attainment Scaling with Children with Sensory Processing Disorder

Abstract

Background: Goal Attainment Scaling (GAS) is advocated as a meaningful outcome measure for parents and clinicians using an Ayres Sensory Integration[®] Intervention (ASI) approach. Although used in several treatment effectiveness studies, reliability of therapist goal writing and scoring has not been established in this area.

Method: Ten occupational therapists and 40 parents of children receiving ASI participated across two clinical sites. The interview therapists and inter-rater therapists wrote GAS goals based on the same goal-setting interviews conducted with parents. Follow-up parent interviews were conducted post-ASI intervention, and the GAS goals were rated by both the interview therapists and inter-rater therapists.

Results: Seventy-eight percent of interview therapist and inter-rater therapist-written goals agreed on content. Intra Class Correlation coefficient of agreement between the two sets of raters was .70 for the total score. Control of bias for establishing and rating the projected level of performance and scaled GAS goals was within an acceptable range.

Conclusion: Findings contribute reliability evidence for use of GAS with children with sensory processing and integration challenges. Two therapists, from different clinical sites who were unfamiliar with the child, identified goal areas, wrote similar GAS goals based on the same parent interview, and scored goals post intervention with good inter-rater reliability.

Comments

The authors report receiving funding from the Wallace Foundation Grant and completing this study while employees of the Spiral Foundation and the Koomar Center.

Keywords

sensory, goal attainment scaling, inter-rater, outcomes, sensory processing, reliability, pediatrics

Cover Page Footnote

We would like to acknowledge the invaluable assistance of Lucy J. Miller, PhD, OTR/L, of the STAR Center, and the therapists of the SPD Foundation and OTA The Koomar Center for their assistance with data collection for this project. We would like to thank the Wallace Foundation and OTA The Koomar Center for funding for this study. Many thanks to Sarah Patane, occupational therapy student, for her assistance with this study.

Credentials Display

Teresa A. May-Benson, ScD, OTR/L, FAOTA; Sarah Schoen, PhD, OTR/L; Alison Teasdale; Jane Koomar, PhD, OTR/L, FAOTA

Copyright transfer agreements are not obtained by The Open Journal of Occupational Therapy (OJOT). Reprint permission for this Applied Research should be obtained from the corresponding author(s). Click here to view our open access statement regarding user rights and distribution of this Applied Research.

The Goal Attainment Scaling (GAS) method was first introduced in the literature by Kiresuk and Sherman in 1968 as an evaluative tool for community mental health programs (Hurn et al., 2006). Over subsequent decades the GAS method was advanced in other therapeutic professions to measure outcomes in settings such as penitentiaries, school-based pediatrics, and physical rehabilitation (King et al., 2000; Lloyd, 1986; Maloney et al., 1978; Ottenbacher & Cusick, 1990; Steenbeek et al., 2007). GAS continues to be a popular evaluative method because it has consistently been found to be sensitive to clinical change, to encourage active collaboration between therapists and clients, and to facilitate client awareness and agreement on realistic expectations of therapeutic progress (Becker et al., 2000; Hurn et al., 2006; King et al., 2000; Ottenbacher & Cusick, 1990; Scott & Haggarty, 1984; Turner-Stokes, 2009; Wright et al., 2005). However, GAS can be a time-consuming process. It requires experienced practitioners to implement as well as an extensive training process for high reliability (Becker et al., 2000; Forbes, 1998; King et al., 2000; Wade, 2009).

Overview of GAS

GAS is a technique for evaluating progress toward individualized goals. It has been used in a variety of professions, such as physical therapy, speech and language therapy, and mental health as well as with a range of clinical populations, such as pediatrics, geriatrics, rehabilitation, and communication disorders. Its strengths include the individualized nature of the goals and its versatility across populations, interventions, and fields, as well as the ability to operationalize outcomes and reflect the client and family's priorities. GAS has the potential to facilitate goal achievement and build teamwork among those engaged in the team process (King et al., 2000; Schlosser, 2004).

While GAS has many strengths, some challenges have been identified. Threats to reliability and validity of the tool exist. For example, unintentional bias in outcomes can occur if goals are written that are too easy to attain. Training ensures that goals are well-written and include the necessary features (e.g., are specific, measurable, realistic, attainable, and time limited with no gaps or overlaps in outcome criteria across scaled levels of performance). It is important that goals are meaningful to the client and the family and can be objectively and consistently scored after a predetermined time frame. Demonstrating consistency in the identification of goal problem areas and creating equidistant scales that reflect outcome difficulty based on present and projected levels of performance are all deemed critical for use of GAS in outcome studies (Ruble et al., 2012).

The GAS Process

There are essentially five steps in the GAS process: (a) establishing the goals in conjunction with the caregiver and family via a qualitative interview process, (b) scaling the goal by operationalizing the expected outcome or level of performance that is predicted to occur after a predefined period and type of intervention and prespecified time of treatment, (c) assigning a weight or rank to the goals based on the therapist's impression of the importance to the family, (d) rating the level of goal achievement at the end of intervention, and (e) examining goals for potential bias at various steps of the GAS process. Goals are written to reflect five levels of projected performance: much less than expected outcome (-2), less than expected outcome (-1), expected outcome (0), more than expected outcome (+1), and much more than expected outcome (+2). Projections for expected level of performance are determined based on current performance levels, the individual's expected rate of change over a given time period, type of intervention, and other individualized factors, such as diagnoses, family situation, etc. When writing GAS goals, the GAS-trained therapist must accurately predict what level of performance the child will most likely demonstrate after a specified period of time given a specific intervention. In addition, when

scaling the goal, the therapist must correctly judge the difficulty in achieving the various levels of expected change so there is equal probability of change between each level and that those levels are equally distributed around the predicted level of performance (i.e., 0). Change is reflected in the outcome level identified during a follow-up parent interview (or child observation) after a preestablished time frame. Although initially assigned an ordinal level when rated and scored for goal achievement, goal sets (at both individual and program levels) can be converted into a standardized *T*-score (mean = 50, standard deviation = 10) because all goals are written to reflect a continuous level of performance (e.g., Level -2 might include performance of a task 10%–20% of the time while Level -1 might include performance at 21%–50% and so on). See Table 1 for a sample GAS goal.

Table 1

Example of a Scaled GAS Goal

Current Level of Performance: “J” becomes very frustrated with fine motor activities. He has a poor grip on the pencil, which decreases his motor control. His difficulties with visual skills are also impacting his willingness to engage in focused tabletop tasks. His mother reports that he will only engage in a fine motor activity for 3–4 min 1–2 times a week.

-2	-1	0	1	2
“J” will engage in a fine motor activity (i.e., drawing, arts and crafts, mazes) for 3–5 min 1–2 times per week.	“J” will engage in a fine motor activity (i.e., drawing, arts and crafts, mazes) for 6–7 min 1–2 times per week.	“J” will engage in a fine motor activity (i.e., drawing, arts and crafts, mazes) for 8–9 min 1–2 times per week.	“J” will engage in a fine motor activity (i.e., drawing, arts and crafts, mazes) for 10–11 min 1–2 times per week.	“J” will engage in a fine motor activity (i.e., drawing, arts and crafts, mazes) for 12 or more min 1–2 times per week.

GAS is a frequently used outcome measure for exploring the effectiveness of interventions. However, psychometric challenges have been reported in the literature that threaten the method’s acceptance by funders and researchers. Four issues of reliability are cited in the literature as important (Stolee et al., 1992): (a) reliability of goal areas identified, (b) reliability in scaling individual goals (e.g., defining the expected outcome) and the scaled levels (setting the difficulty range of scaled goals), (c) reliability in rank ordering the importance of goals, and (d) reliability in scoring the outcome (e.g., follow-up goal score). These issues are addressed individually below.

Goal Content Agreement

Reliability in identification of goal areas is reflected in the content agreement of goal problem areas across clinicians in the construction of goals. Precisely written goals are idiosyncratic to the client or patient and, as such, are usually based on assessment or interview findings. When written correctly, goals should be specific, objective, and measurable. Investigations of interrater agreement in the identification of GAS goal areas has not been widely studied. Results are reportedly mixed across studies. While goal construction may not be a requirement of reliability for GAS (Schlosser, 2004; Smith, 1994), it is recommended that goals emerge from similar problem areas.

Goal Weighting Agreement

Weighting is the process of rank ordering goals to quantify the relative importance of the selected goals. Although this might be an attractive feature of GAS, particularly for clinicians, the literature suggests reliability for such judgments is potentially poor and acknowledges that weighting the goals complicates the calculation of summary *T*-scores at follow-up. If clinicians and researchers chose

to weight client goals, then it is recommended that interrater reliability in the assignment of weights be demonstrated (Marson et al., 2009).

Goal Outcome Rating Agreement

Interrater reliability of GAS scoring at follow-up has been a primary area of focus in the literature (Marson et al., 2009). Rating goal achievement means determining the level of performance and goal outcome achievement at follow-up in the goals that were set, after the prespecified time of treatment or intervention. A recent systematic review of 37 studies indicated that 74% of the papers reviewed showed evidence of interrater reliability (Shankar et al., 2020). Since the focus of Shankar et al. (2020) was on validity evidence, reliability values were obtained from other literature that report agreement ranges from .82 to .98 (King et al., 2000; Ruble et al., 2012; Schlosser, 2004; Steenbeek et al., 2010). It is recommended, however, that the reliability of GAS scores be calculated in each new application (Schlosser, 2004). None of the studies included children with sensory processing and integration challenges.

Control of Bias

Therapist bias in goal construction and goal rating has been identified as a potential weakness in the reliability of the GAS methodology (King et al., 2000). Concerns have been expressed over inconsistencies among therapists when writing individualized goals, setting predicted outcomes, and measuring outcomes (Schlosser, 2004). Further concerns have been expressed about goal writers setting predicted levels of performance at levels that are too easy or too hard to achieve and setting the scaled goal benchmarks inconsistently in terms of difficulty of attainment (Donnelly & Carswell, 2002; Ruble et al., 2012). Bias in these areas can contribute to unreliable outcome measurement. Therefore, use of GAS must include monitoring these potential sources of bias to ensure goals and outcomes reliably reflect goal attainment. This may be studied by examination of predicted outcomes for difficulty level (in relation to present level of performance) as well as the manner in which the goals are scaled. Bias is controlled for by clinicians accurately predicting realistic and clinically meaningful expected outcome levels when setting the goals. According to Sherman (1994), outcome *T*-scores would be expected to converge at 50 based on the theorized distribution of scaled goals (e.g., 0 level should be the expected outcome), with equidistant intervals (in terms of difficulty in achievement) set between each level (standard deviation of 10). This means that expected outcomes should not be underestimated or overestimated. Clinicians should not set expectations of goal attainment too low just to increase the likelihood of success for the child, and they should not set expectations too high such that the child cannot achieve goal attainment. In addition, bias is further reduced by ensuring that the scaled levels are continuous and clearly operationally defined across the measurement parameters and selected and benchmark descriptions. Although the literature discusses the importance of properly setting the predicted outcome and scaling goals, examination of bias in GAS goals setting and rating has rarely been reported in the literature (Schlosser, 2004). For more details on the GAS methodology, readers are referred to Kiresuk, Smith, and Cardillo's (1994) text.

Application of GAS

Both Donnelly and Carswell (2002) and Hurn et al. (2006) reported strong interrater reliability across adult rehabilitation practice settings. Reliability has been found to significantly increase when therapists are consistently and extensively trained in GAS techniques and given adequate practice and instruction (Steenbeek et al., 2007). Training in goal writing can facilitate the construction of high-quality GAS scales, minimize the risk of bias, and ensures that GAS goal descriptions are clear,

measurable, and follow the criteria outlined in the literature (Krasny-Pacini et al., 2016). Reliability is also increased when goals are set and rated by a therapist other than the treating therapist and further increased with multiple rating therapists (Palisano, 1993; Steenbeek et al., 2007). The reliability of weighting of goals was examined in one study that found that goal weighting is reliable in similarly trained therapists (Marson et al., 2009). The majority of these studies use only experienced GAS-trained therapists to measure validity and reliability of the scales (Palisano, 1993; Steenbeek et al., 2007), which may not represent the average therapist who uses GAS in clinical occupational therapy practice.

GAS has been advocated for use with individuals with sensory processing and integration challenges. Mailloux et al. (2007) reported that GAS is a meaningful measure for parents and clinicians using an Ayres Sensory Integration[®] (ASI) approach and advocated for further examination of this goal setting methodology (Mailloux et al. 2007). Subsequently, GAS has been a primary outcome measure in several treatment studies (Miller et al., 2007; Pfeiffer et al., 2011; Schaaf et al., 2012; Schaaf et al., 2014) and has been identified as a primary outcome establishing ASI as an evidence-based practice (Schoen et al., 2019). In intervention studies, as in clinical practice, it is not practical or desired for interrater therapists to be present during both goal setting and follow-up interview sessions with parents. As a result, interviews are typically video recorded and additional ratings are conducted from these recordings. This makes it necessary to ensure that goal data obtained from live interviews is consistent with that obtained from video-recorded interviews.

The purpose of this study was to examine the interrater reliability between GAS-trained occupational therapists in identifying, writing, and following up rating of GAS goals for children with sensory processing disorders in both live interview and video-recorded interview settings. Further, control of bias in goal writing was examined.

The first research question was: What is the inter-therapist reliability in GAS goal setting, rank ordering and weighting of goals for importance to the family, and follow-up rating of GAS-trained therapists? To address the research question, the following hypotheses were examined:

1. Following a parent goal-setting meeting, GAS-trained therapists will demonstrate significant interrater agreement on goal content of intervention problem areas.
2. Following a parent goal-setting meeting, GAS-trained therapists will demonstrate significant interrater agreement on rank order and weighting of goal importance.
3. GAS-trained therapists will demonstrate adequate interrater reliability in their determination of a child's level of performance on GAS goals in follow-up scoring after intervention.

The second research question was: Do goals written by GAS-trained therapists demonstrate adequate control of bias in setting expected outcomes and range of scaled outcomes from live and video-recorded interviews? To address the research question, the following hypotheses were addressed:

1. GAS-trained therapists will demonstrate adequate control of bias of goal construction as evidenced by a mean *T*-score between 40 and 60 in setting expected outcomes for GAS goals in both live and video-recorded settings.
2. GAS-trained therapists will demonstrate adequate control of scaling bias in setting the difficulty range of the scaled goals as evidenced by a mean standard deviation *T*-score range of 8 to 12 for the range of predicted scaled outcomes for GAS goals in both live and video-recorded settings.

Method

This study was approved by the Spiral Foundation Institutional Review Board, protocol #1001. Informed consent was obtained from all participants. This study was carried out in accordance with the

recommendations of the Office of Human Rights Protection. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Participants

Therapists

This study included 10 occupational therapists from two private occupational therapy clinics specializing in ASI: five from a site in the Northeast US (Site 1) and five from a site in the Western US (Site 2). All therapists had a minimum of 3-years of experience in sensory integration intervention. Because GAS is a complex goal writing methodology that involves goal setting strategies that are not typically used by therapists, extensive training in the theory, methodology, and practical application of GAS was required prior to study implementation. The GAS methodology was manualized by the first author and a training program was developed for this study based on the first author's extensive research knowledge and clinical expertise using this model and has since been provided as a continuing education course internationally. (The GAS manual and training materials are available from the first author.) Therefore, therapist-participants at both sites completed a 24-hour GAS training program, provided by the first author, which involved 1 day of in-class training and 18 hr of subsequent practical application of the GAS process with peer feedback. Two therapists from Site 2 were hired after the initial trainings and were trained by the second author, who had attended the initial training. Individual therapists served as interview therapists who conducted both live parent goal-setting and follow-up goal rating interviews in their own clinical setting and as interrater therapists who reviewed video recordings of parent goal-setting and follow-up goal rating interviews from the second clinical setting. In the live interviews, the same therapist completed both goal-setting and follow-up interviews per family. Interrater therapists reviewed either goal setting or follow-up video-recorded interviews.

Parents

Parent-participants were recruited as a convenience sample from clients beginning services at each clinical site. Forty parents of children with sensory processing disorder between 3 and 11 years of age and who were receiving ASI occupational therapy services at the study facilities participated. Parents of 19 children were recruited from Site 1 and parents of 21 children from Site 2. An occupational therapist experienced in sensory integration assessment previously identified the children with sensory processing and integration challenges following a comprehensive evaluation using standardized measures of sensory-motor function (e.g., Sensory Integration and Praxis Tests, the Miller Function and Participation Scales, or the Bruininks-Oseretsky Test of Motor Proficiency) and clinical observation and parent report as part of routine clinical practice at each site. The children had average intelligence as identified by age level performance in school and no major medical or psychological diagnoses, with the exception of attention deficit disorder, non-verbal learning disability, or anxiety disorder. No other specific inclusion or exclusion criteria were applied since the children were not participants in the study.

Measures

GAS methodology was used. GAS goals identified and written by therapists as a result of a parent goal-setting meeting (or a video recording of the goal-setting meeting) were used for this study. A set of five GAS goals were identified and written. Goals were written to reflect functional performance of activities meaningful to the family in a nonclinical setting in a way such that progress could be potentially observed and reported on by the parents. Identification of goals was based on an in-person parent goal-setting meeting for the child of each of the 40 parent-participants, which resulted in 40 GAS

goals sets (each set with five goals) and 200 individual goals. Each specific goal was scaled into a range of projected outcomes that reflected varying levels of change that could occur over a specified intervention period. Thus, each goal was “scaled” into five levels of potential performance from -2 to +2, as previously noted. After a projected 20 sessions (actual number of sessions varied from 18 to 20), the child’s post intervention level of performance on each goal was determined by the therapists based on an interview (or video recording of the interview) with the parents.

Procedures

This study was conducted at two clinical sites. Goal setting interviews and intervention were completed in the context of the routine clinical practice of each clinical site. Clinical procedures varied slightly between the clinical sites. At Site 1, the clients engaged in a phone intake interview with a clinician from the clinic, completed an occupational therapy evaluation with parent meeting, and then completed the GAS goal setting interview to establish goals for intervention. At Site 2, the clients completed an in-person intake interview and a comprehensive occupational therapy evaluation followed by a parent meeting for evaluation feedback and the GAS goal setting interview. In both settings, the same therapist completed the live follow-up interviews for goal rating. Follow-up goal rating was based only on information provided by the parents.

Informed consent was obtained from all parent-participants prior to enrollment in the study. A goal-setting interview with the parent-participants lasting approximately 1 hr was conducted and videotaped by a GAS therapist at the parent’s clinical facility. The interview therapist then identified and wrote five specific scaled GAS goals that reflected functioning in areas of importance to the families of the child that were reported during the goal-setting session. Each child was scheduled to receive 20 1-hr sessions of occupational therapy intervention using an ASI approach. After the intervention period, a GAS therapist and the parent-participant met for a 1-hr follow-up meeting to review the child’s performance on the GAS goals. All interviews were qualitative interviews involving open-ended questions. The goal of both the pre and post intervention interviews was to elicit information about the child’s functional performance without specifically guiding or biasing parent report. Methodology for performing the interviews was provided during the GAS training received by each therapist. See Table 2 for examples of interview questions.

Table 2

Example Questions for Semi-Structured Qualitative Pre and Post Intervention GAS Interview.

Pre intervention goal setting interview questions.

- Tell me about your child. What are his/her strengths, his/her weaknesses?
 - What has led you to seek services for your child?
 - What concerns you most about your child? Tell me more specifically about
 - Tell me about a typical week day . . . a typical weekend day.
 - Review the child’s evaluation and ask questions regarding areas of specific difficulty. i.e., I notice that _____ seems to be hard for him/her. Can you tell me more about that? OR, Tell me more specifically about (each specific sensory area identified as problematic from the evaluation).
 - Our evaluations showed some difficulties/delays with Is this something that has been of concern to you?
 - What are some goals you have for your child, in the next 6 months or so? (time frame may be variable)
 - Looking ahead to the end of this 20-session intervention period, what are reasonable goals that you hoping your child can accomplish?
-

Post intervention follow-up goal rating interview questions

- Tell me about how things are going these days
 - Tell me more about.... (elicit more information on topics parent brings up basic on previous question.)
 - (If parents do not independently bring up information about GAS goal areas ask very broad questions about topic to elicit information on goal performance.) I notice that _____ had been having some challenges in _____ (broad area, e.g. dressing, eating, mealtime, etc.). Can you tell me more about that?
-

Based on information obtained during the post intervention interview meeting, the interview therapist and parent-participant independently rated the child's level of performance on the GAS goals. Interrater therapists from one clinical setting viewed a video recording of the interviews, independently wrote a second set of goals for the child, and ranked and rated the first set of goals from the other clinical setting. Data on the reliability between the interview therapist and the interrater therapist are reported in this study. Data on reliability between parents and therapists will be reported in another publication.

Twenty (10 from each site) pre intervention interview videos and nine post intervention videos from Site 1 and six from Site 2 were randomly selected for interrater reliability checks. Last, there were four participants missing follow-up data from Site 2, which resulted in slightly lower numbers than expected available for analysis of outcome rating agreement. The research assistant was responsible for completing the analyses of content agreement, rank order of importance agreement, goal achievement outcome agreement, and control of bias.

Data Analysis

Data from goals written by the therapist who conducted the parent interview and an interrater therapist from the other clinical site who reviewed a video recording of the interview were examined. None of the interview or interrater therapists treated the children whose parents were interviewed. Forty GAS goal sets (five individual goals per set), including 200 individual goals, were established by the interview therapists from both sites prior to intervention; however, only a random sample of goals were rated for some analyses. Thus, the number of goals (e.g., individual goals) or goal sets (e.g., set of five individual goals) varied by analysis.

Analysis of content of 50 randomly selected individual goals, out of 200 available goals, was completed by the research assistant (the third author). Key words in the interview therapist's goals, which reflected goal content areas of functional performance, were identified. Key words were then similarly identified in the interrater goals written for that client. Using the *Occupational Therapy Practice Framework* (OTPF-3; American Occupational Therapy Association, 2014) as a guide, goals were deemed an exact match if outcomes reflected by the keywords were the same between the two sets of goals (e.g., if the interview goal related to the child's ability to don a shirt and the interrater goal related to donning a shirt). Goals were deemed to be a partial match if the area of functional performance (as found in the OTPF-3) was the same but the specific task was different, (e.g., the interview goal related to the child's ability to don a shirt, a dressing ADL, while the alternate goal involved donning pants, a similar dressing ADL). Goals not of the same OTPF-3 categories were considered non-matches. Frequency counts of exact and partial interrater agreement on goal setting content areas were completed.

Goals were also examined based on agreement relative to rank order of importance as well as ratings of goal achievement outcomes. Examination of ability of two therapists to similarly weight GAS

goals by importance was completed using a Kendall's Tau-b correlation coefficient analyses (a non-parametric measure of relationships between columns of ranked data). Interrater agreement on ratings of goal achievement outcomes was completed by calculation of intra-class correlations between raters. Control of bias for goal setting was examined by calculating *T*-score outcome mean and standard deviations according to established GAS protocols for the interview and interrater groups and comparing those scores to the control of bias charts provided by Kiresuk et al. (1994). Where possible differences between clinical sites were examined.

Results

Goal Content Agreement

The 50 paired goals were compared between the interview and interrater therapists. Exact agreement on goal content was found for 63% of 50 individual goal pairs and partial agreement for an additional 15% of goals. Thus, overall agreement across sites was found for 78% of goals, indicating there is high interrater agreement between two therapists in identification of content of intervention goals based on parent interview. Examination of data by site found that 88% of interrater goals written by therapists at Site 1 agreed with the original goal content generated by the therapists at Site 2, while 68% of interrater goals written by therapists at Site 2 agreed with the original goal content generated by the therapists at Site 1.

Goal Weighting Agreement

Interrater comparisons between the interview therapist and interrater therapist were completed for goal weighting (rank order of importance) on $n = 20$ paired GAS goal sets (each set consisted of five goals). Kendall's Tau-b Correlation Coefficient ($Tau-b = -.281 - .301, p = .106 - .916$) analysis (Puka, 2011) found there was no significant correlation of weighted scores between the interview therapist and interrater therapist, indicating the two therapists did not weight the five goals in each GAS goal set the same.

Goal Outcome Rating Agreement

Fifteen GAS goal sets were examined for post intervention outcome goal rating agreement, $n = 9$ from Site 1 and $n = 6$ from Site 2. Ratings from the interview therapist were compared to ratings from the interrater therapist of the video interview of the parent. Intra-class correlation (2-way mixed-people effects fixed and measure effects random) for absolute agreement was adequate for total score ($ICC = .702$). Thus, the interview therapists and interrater therapists rating the same interview from video were in agreement.

Examination of site data, however, found that ratings of goal outcome agreement between the video raters and the interview therapist ratings were stronger for goals written at Site 1 than those at Site 2 ($ICC = .811$). Ratings for outcome agreement for goals written at Site 2 were only available for six sets of goals. Agreement between the interview therapists at Site 2 and video raters at Site 1 was poor ($ICC = .412$).

Control of Bias in Writing Goals from Live Interviews

The ability of the interview therapists to control bias in difficulty in achieving goals when writing goals and to accurately establish the expected level of performance when obtaining information from a live interview was examined for $n = 33$ GAS goal sets resulting in scores for 164 goals across 10 therapists. Determination of an acceptable level of difficulty was based on *T*-score range as described by Sherman (1994). The mean of a set of goals should be $T = 50$ with a range of $T = 46.5$ to $T = 53.5$. Bias of outcome ratings by the interview therapist found the mean outcome of $n = 164$ goals to be excellent

with a mean T -score of 50.0 and a standard deviation of 10.65. These scores were compared to the control of bias charts provided by Kiresuk et al. (1994). Thus, results indicate that, overall, the interview therapists set the expected level of performance at an acceptable level of difficulty.

For GAS goals it is also important to examine bias in setting the difficulty of achieving a range of the scaled goals to ensure reliable goal rating. This is reflected in the average standard deviation of the T -score of a set of GAS goals. The interview therapists had an excellent overall mean T -score standard deviation of 10.65. According to Sherman (1994) the mean standard deviation of T -scores should be 10 with an acceptable range of 8 – 12. Thus, the interview therapists tended to write and rate their scaled goals appropriately.

Examination of control of bias by site using an independent samples t -test found a marginally significant difference ($t(31) = 2.040, p = .050$) between sites for bias in setting the difficulty of achieving the expected level of performance. However, goals written and rated by the therapists at Site 1 ($n = 16$ sets) were within the expected limits with a goal outcome t -score $m = 46.3$ with an expected range of 45.8 – 54.2. Goals written and rated by the therapists at Site 2 (17 sets) were also within expected limits with a t -score $m = 53.5$ with an expected range of 45.9 – 54.1. Bias in setting the range of scaled goals was within expected limits for both sites with Site 1 $sd = 11.72$ and Site 2 $sd = 8.44$. Thus, although the mean T -scores of both sites were significantly different from each other, they both fell within the accepted range of scores.

Control of Bias in Rating Goals from Video-Recorded Interviews

Control of bias in rating goals at follow-up using video-recorded interviews was also examined. Overall, therapists rating $n = 15$ goal sets from video interviews had a mean T -score of 44.2, which is outside the expected range of 45.7 – 54.3. The standard deviation of these ratings was 10.02 and was within expected limits.

Examination of bias in rating goals from video interviews by site using an independent samples t -test found no significant difference ($t(13) = 1.259, p = .230$) between sites. Goals written by the therapists at Site 1 but rated by the therapists at Site 2 ($n = 9$ sets, 45 individual goals) had a T -score $m = 41.6, sd = 8.50$, which was outside the expected range of 45.1 – 54.9 for the mean. Goals written by the therapists at Site 2 and rated by the therapists at Site 1 ($n = 6$ sets, 29 individual goals) fell within acceptable limits with a t -score $m = 48.1, sd = 11.62$ with an expected range of 44.4 – 55.6 for the mean.

Discussion

With the increase in popularity of GAS in clinical and research practice, concern has been expressed regarding the application of this methodology (Harpster et al., 2018). The psychometric properties of GAS have largely been explored in pediatric rehabilitation settings (Harpster et al., 2018; Krasny-Pacini et al., 2016; Steenbeek et al., 2010; Steenbeek et al., 2011). This study extends the application of GAS to children with sensory processing and integration challenges and contributes to the psychometric validation of how GAS is applied in this population.

Goal Content Agreement

Research ensures that GAS is being used with a high degree of methodological rigor (Harpster et al., 2018). This study adds to the already existing literature by demonstrating interrater reliability for the construction of GAS goals. Two therapists, from different clinical sites who were unfamiliar with a child, were able to identify goal areas and write similar GAS goals based on the same parent interview, thus demonstrating good interrater reliability between clinicians. This provides preliminary evidence

that similarly trained therapists are potentially able to consistently identify parent priority goal area content from an interview when using the GAS for measuring intervention outcomes.

Goal Weighting Agreement

This study also sheds light on the question of clinical importance of weighting parent goals since no significant agreement was found between the interview therapist and interrater therapist for weighting the importance of goals. This finding is consistent with Kiresuk et al.'s (1994) assertion that weighting of goals may not be clinically relevant. It appears likely that all goals are important and the small differences in priority among the goals are not clinically meaningful to families, thus making rank order weighting difficult. Although one study found acceptable interrater reliability for assigning weights to GAS goals (Marson et al., 2009), methodological differences likely affected the results (e.g., they used a hypothetical, predigested case summary and student raters enrolled in a social work course). Regardless, this study suggests that rank ordering the importance of parent identified goals may not be necessary when therapists or researchers use GAS as an outcome measure. While there is likely to be some implicit weighting with goal writing, clinical experience recommends having different categories (e.g., of body function, activity performance, or participation) to prevent writing two goals on the same need area.

Goal Outcome Rating Agreement

In addition, this study explored interrater reliability between therapists when scoring GAS goals. The ability to have agreement on ratings of goal outcomes between therapists is particularly important if GAS is to be used as an outcome measure in research or clinical practice. In general, this study found that interrater agreement for rating level of outcome goal achievement was good. This suggests that following a parent interview, a therapist who is unfamiliar with a child is able to rate the child's performance from a video recording of the interview the same as a therapist who conducts the interview in-person. This is similar to results obtained by Ruble et al. (2012), who also found consistency in scores gathered on outcomes though videotape. Thus, for research studies or clinical practice, video-recorded interviews may be completed by one person and outcome ratings may be reliably completed by a different therapist.

It is likely that these results were facilitated by the training therapists received in this study. As recommended in the literature, the therapists were trained in the interview process and writing GAS goals. The participant-therapists completed an extensive training provided by the first author or by a therapist who had attended the training. As this study showed, and previous research indicates, training is key to improving clinical effectiveness and rigor of GAS as an outcome measure (Krasny-Pacini et al., 2016). Contributing to the success of this training was the fact that the team who was trained had experience with this specific population and knowledge of what could be attained in a given time frame.

Control of Bias

When examining control of bias in rating goals, it was found that the therapists who rated goals from video interviews demonstrated a mean rating that fell just outside of the expected range for acceptable control of bias. Specifically, this was noted for the therapists at Site 2 who scored goals written by the therapists at Site 1. This suggests the possibility of some unreliability in rating goals from video recording. Several hypotheses for this finding are proposed. It is possible that the therapists who rate goal achievement from video recordings employ stricter criteria than the therapists that rate from live interviews. It is also possible that training differences or clinical procedures at the two sites resulted in slightly different rating criteria across sites. Further research is needed to answer this question.

Since GAS goals are subjectively written and rated, to have accurate and reliable goals it is important to control for bias in both goal writing and goal outcome ratings. This study found that the therapists who conducted the parent interviews demonstrated excellent control of bias in setting the expected level of performance as well as in scaling GAS goals. Although there was a marginally significant difference between sites, goals written and rated by the therapists from each site were within the acceptable range of scores. It is interesting, however, that the mean *T*-score of Site 1 ratings was at the low end of the range while mean ratings of Site 2 were at the high end of the acceptable range. Controlling bias in this way ensures that the outcomes reported using this methodology are valid and not because of goal writers making outcomes too easy to achieve, thus potentially biasing outcomes to positive results.

Limitations

This study presents with several limitations that may have influenced the results. While randomized goals were selected for analysis of goal content and interrater analysis, all available goals were used for other analyses. Subsequently, the sample size in some analyses were smaller than anticipated, which may have resulted in more variability in findings. Although all therapists received GAS training, unforeseen staffing problems at one site resulted in some differences in training for several staff that may have impacted the way that interviews were conducted by those therapists. The results may have been stronger with more consistent training. These differences may have impacted the control of bias findings. In addition, while both sites followed GAS guidelines, individualized interview and goal writing styles of the two sites were somewhat different, which may have contributed to differences in outcome ratings. Ensuring more consistency in therapist experience, interview style, goal writing, and rating procedures across sites would be helpful in improving reliability among raters.

It is important to recognize that this was not an intervention study and the authors do not claim that the ASI intervention was responsible for the progress observed. There was no attempt to control for the intervention that was provided other than to ensure that the ASI intervention was a central component of the intervention program. Since both clinics routinely use other ancillary interventions, it is possible that the children's progress may have been impacted by a range of factors. This issue would be addressed in a well-controlled intervention study. In addition, the goal of this study was to examine reliability issues potentially inherent in the GAS process using rigorous methodology as recommended by Palisano (1993) and Steenbeek et al. (2007). As such, while use of a different therapist from the treating therapist for goal setting and follow-up goal rating may limit therapeutic rapport with parents in a clinical setting, this methodology was necessary for this study. In a clinic setting, this procedure may need to be examined and ways to ensure maintenance of the therapeutic value of parent-therapist meetings should be considered. Finally, goal achievement in this study was based on parent interview rather than observation. This procedure is not uncommon when using the GAS process, as family-centered goals are often best evaluated by the parent. However, future intervention studies might alter this process by having an unbiased observer score goal achievement.

Clinical Application of Findings

Consistent with a recent systematic review that emphasizes the use of GAS for efficacy and program evaluation (Harpster et al., 2018), this study presents very promising preliminary reliability evidence for the use of GAS as a meaningful outcome of ASI occupational therapy intervention. Despite criticism in the literature about the potential bias of GAS (Weitlauf et al., 2017), this study is consistent with findings of previous studies, that goals can be objectively scored and rated similarly by two

independent observers (Ruble et al., 2012; Steenbeek et al., 2010). In addition, GAS should be considered complementary to standardized instruments since it captures individual-specific outcomes (Steenbeek et al., 2011).

This study further highlights that implementation of GAS in clinical practice presents with several strengths and limitations. The strengths of GAS include the ability to identify individualized intervention goals that are meaningful to the client and family (Schaaf et al., 2018). GAS has the ability to capture small increments of change in functional areas of performance that may not be possible on standardized measures. In addition, the GAS process supports parents as active participants in identifying and establishing their child's goals. As GAS scores are standard scores, outcome results from multiple clients can be combined to provide information on program and intervention effectiveness (Kiresuk et al., 1994). This study also supports previous studies that challenge the need for weighting of individual goals. Challenges of implementation of GAS in clinical settings include factors such as the need for training and experience for clinicians to be reliable goal setters and raters. The GAS process is also time-consuming in terms of conducting parent interviews, setting scaled goals, and conducting follow-up goal rating. There is a need to monitor goal setting and ratings on a regular basis to ensure that bias is not present in either situation. In spite of these challenges, GAS may be recommended for research and clinical use when examining outcomes of the ASI intervention.

References

- American Occupational Therapy Association. (2014). Occupational therapy practice framework: Domain and process (3rd ed.). *American Journal of Occupational Therapy*, 68(Suppl. 1), S1–S48. <http://dx.doi.org/10.5014/ajot.2014.682006>
- Becker, H., Stuijbergen, A., Rogers, S., & Timmerman, G. (2000). Goal attainment scaling to measure individual change in intervention studies. *Nursing Research*, 49(3), 176–180. <http://dx.doi.org/10.1097/00006199-200005000-00011>
- Donnelly, C., & Carswell, A. (2002). Individualized outcome measures: A review of the literature. *Canadian Journal of Occupational Therapy*, 69(2), 84–94. <https://doi.org/10.1177/000841740206900204>
- Forbes, D. A. (1998). Goal attainment scaling: A responsive measure of client outcomes. *Journal of Gerontological Nursing*, 24(12), 34–40. <https://doi.org/10.3928/0098-9134-19981201-08>
- Harpster, K., Sheehan, A., Foster, E. A., Leffler, E., Schwab, S. M., & Angeli, J. M. (2018). The methodological application of goal attainment scaling in pediatric rehabilitation research: A systematic review. *Disability and Rehabilitation*, 1–10. <https://doi.org/10.1080/09638288.2018.1474952>
- Hurn, J., Kneebone, I., & Cropley, M. (2006). Goal setting as an outcome measure: A systematic review. *Clinical Rehabilitation*, 20(9), 756–772. <https://doi.org/10.1177/0269215506070793>
- King, G. A., McDougall, J., Palisano, R. J., Gritzan, J., & Tucker, M. A. (2000). Goal attainment scaling. *Physical & Occupational Therapy in Pediatrics*, 19(2), 31–52. https://doi.org/10.1080/J006v19n02_03
- Kiresuk, T. J., Smith, A., & Cardillo, J. E. (1994). *Goal attainment scaling: Applications, theory and measurement*. Lawrence Erlbaum Associates.
- Krasny-Pacini, A., Evans, J., Sohlberg, M. M., & Chevignard, M. (2016). Proposed criteria for appraising goal attainment scales used as outcome measures in rehabilitation research. *Archives of Physical Medicine and Rehabilitation*, 97(1), 157–170. <https://doi.org/10.1016/j.apmr.2015.08.424>
- Lloyd, C. (1986). The process of goal setting using goal attainment scaling in a therapeutic community. *Occupational Therapy in Mental Health*, 6(3), 19–30. https://doi.org/10.1300/J004v06n03_02
- Mailloux, Z., May-Benson, T. A., Summers, C. A., Miller, L. J., Brett-Green, B., Burke, J. P., Cohn, E. S., Koomar, J. A., Parham, D. L., Roley, S. S., Schaaf, R. C., & Schoen, S. A. (2007). Goal attainment scaling as a measure of meaningful outcomes for children with sensory integration disorders. *American Journal of Occupational Therapy*, 61(2), 254–259. <http://dx.doi.org/10.5014/ajot.61.2.254>
- Maloney, F. P., Mirrett, P., Brooks, C., & Johannes, K. (1978). Use of the Goal Attainment Scale in the treatment and ongoing evaluation of neurologically handicapped children. *American Journal of Occupational Therapy*, 32(8), 505–510.
- Marson, S. M., Wei, G., & Wasserman, D. (2009). A reliability analysis of Goal Attainment Scaling (GAS) weights. *American Journal of Evaluation*, 30, 203. <https://doi.org/10.1177/1098214009334676>
- Miller, L. J., Schoen, S. A., James, K., & Schaaf, R. C. (2007). Lessons learned: A pilot study on occupational therapy effectiveness for children with sensory modulation disorder. *American*

- Journal of Occupational Therapy*, 61(2), 161–169. <https://doi.org/10.5014/ajot.61.2.161>
- Ottenbacher, K. J., & Cusick, A. (1990). Goal attainment scaling as a method of clinical service evaluation. *American Journal of Occupational Therapy*, 44(6), 519–525. <https://doi.org/10.5014/ajot.44.6.519>
- Palisano, R. J. (1993). Validity of goal attainment scaling in infants with motor delays. *Physical Therapy*, 73(10), 651–658. <https://doi.org/10.1093/ptj/73.10.651>
- Pfeiffer, B. A., Koenig, K., Kinnealey, M., Sheppard, M., & Henderson, L. (2011). Effectiveness of sensory integration interventions in children with autism spectrum disorders: A pilot study. *American Journal of Occupational Therapy*, 65(1), 76–85. <https://doi.org/10.5014/ajot.2011.09205>
- Puka, L. (2011). Kendall's Tau. In M. Lovric (Ed.), *International encyclopedia of statistical science*. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-04898-2>
- Ruble, L., McGrew, J. H., & Toland, M. D. (2012). Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-012-1446-7>
- Schaaf, R. C., Dumont, R. L., Arbesman, M., & May-Benson, T. A. (2018). Efficacy of occupational therapy using Ayres Sensory Integration®: A systematic review. *American Journal of Occupational Therapy*, 72(1), 7201190010p7201190011–7201190010p7201190010. <https://doi.org/10.5014/ajot.2018.028431>
- Schaaf, R. C., Benevides, T., Mailloux, Z., Faller, P., Hunt, J., van, H. E., Freeman, R., Leiby, B., Sendecki, J., & Kelly, D. (2014). An intervention for sensory difficulties in children with autism: A randomized trial. *Journal of Autism and Developmental Disorders*, 44(7), 1493–1506. <https://doi.org/10.1007/s10803-013-1983-8>
- Schaaf, R. C., Benevides, T. W., Kelly, D., & Mailloux, Z. (2012). Occupational therapy and sensory integration for children with autism: A feasibility, safety, acceptability and fidelity study. *Autism*. <https://doi.org/10.1177/1362361311435157>
- Schlösser, R. W. (2004). Goal attainment scaling as a clinical measurement technique in communication disorders: A critical review. *Journal of Communication Disorders*, 37(3), 217–239. <https://doi.org/10.1016/j.jcomdis.2003.09.003>
- Schoen, S. A., Lane, S. J., Schaaf, R. C., Mailloux, Z., Parham, L. D., Roley, S. S., & May-Benson, T. (2019). Ayres Sensory Integration® meets criteria for an evidence-based practice: A response to Stevenson [2019]. *Autism Research*. <https://doi.org/10.1002/aur.2164>
- Scott, A. H., & Haggarty, E. J. (1984). Structuring goals via goal attainment scaling in occupational therapy groups in a partial hospitalization setting. *Occupational Therapy in Mental Health*, 4(2), 39–58. https://doi.org/10.1300/J004v04n02_04
- Shankar, S., Marshall, S. K., & Zumbo, B. D. (2020). A systematic review of validation practices for the goal attainment scaling measure. *Journal of Psychoeducational Assessment*, 38(2), 236–255. <https://doi.org/10.1177/0734282919840948>
- Sherman, R. E. (1994). Keeping follow-up guides on target. In T. J. Kiresuk, A. Smith, & J. E. Cardillo (Eds.), *Goal attainment scaling: Applications, theory, and measurement* (pp. 279–282). Lawrence Erlbaum Associates.
- Smith, A. (1994). Introduction and overview. In T. J. Kiresuk, A. Smith, & J. E. Cardillo (Eds.), *Goal attainment scaling: Applications, theory, and measurement* (pp. 1–14). Lawrence Erlbaum Associates.
- Steenbeek, D., Gorter, J. W., Ketelaar, M., Galama, K., & Lindeman, E. (2011). Responsiveness of goal attainment scaling in comparison to two standardized measures in outcome evaluation of children with cerebral palsy. *Clinical Rehabilitation*, 25(12), 1128–1139. <https://doi.org/10.1177/0269215511407220>
- Steenbeek, D., Ketelaar, M., Lindeman, E., Galama, K., & Gorter, J. W. (2010). Interrater reliability of goal attainment scaling in rehabilitation of children with cerebral palsy. *Archives of Physical Medicine and Rehabilitation*, 91(3), 429–435. <https://doi.org/10.1016/j.apmr.2009.10.013>
- Steenbeek, D., Ketelaar, M., Galama, K., & Gorter, J. W. (2007). Goal attainment scaling in paediatric rehabilitation: A critical review of the literature. *Developmental Medicine and Child Neurology*, 49(7), 550–556. <https://doi.org/10.1111/j.1469-8749.2007.00550.x>
- Stolee, P., Rockwood, K., Fox, R. A., & Streiner, D. L. (1992). The use of goal attainment scaling in a geriatric care setting. *Journal of the American Geriatric Society*, 40(6), 574–578. <https://doi.org/10.1111/j.1532-5415.1992.tb02105.x>
- Turner-Stokes, L. (2009). Goal attainment scaling (GAS) in rehabilitation: A practical guide. *Clinical Rehabilitation*, 23(4), 362–370. <https://doi.org/10.1177/0269215508101742>
- Wade, D. T. (2009). Goal setting in rehabilitation: An overview of what, why and how. *Clinical Rehabilitation*, 23(4), 291–295. <https://doi.org/10.1177/0269215509103551>
- Weitlauf, A. S., Sathe, N., McPheeters, M. L., & Warren, Z. E. (2017). Interventions targeting sensory challenges in autism spectrum disorder: A systematic review. *Pediatrics*, 139(6). <https://doi.org/10.1542/peds.2017-0347>
- Wright, F. V., Boschen, K., & Jutai, J. (2005). Exploring the comparative responsiveness of a core set of outcome measures in a school-based conductive education programme. *Child Care Health and Development*, 31(3), 291–302. <https://doi.org/10.1111/j.1365-2214.2005.00511.x>