



Published in final edited form as:

J Autism Dev Disord. 2022 August ; 52(8): 3344–3352. doi:10.1007/s10803-021-05213-3.

Goal Attainment Scaling: An Idiographic Measure Sensitive to Parent and Teacher Report of IEP Goal Outcome Assessment for Students with ASD

Lisa Ruble¹, John McGrew², Brittany Dale¹, Madison Yee³

¹Department of Special Education, Ball State University

²Department of Psychology, Indiana University-Purdue University

³Department of Educational, School, and Counseling Psychology, University of Kentucky

Abstract

Young students with ASD have instructional needs in social, communication, and learning skills that should be reflected in their Individualized Education Program (IEP). Research suggests that many of these goal areas present a challenge for special educators because of problems with measurability. The current study utilized an idiographic approach called Goal Attainment Scaling (GAS) for measuring IEP progress of individualized skills. Cross-sectional analysis of the associations between GAS ratings from an independent observer were correlated against teacher and parent ratings of IEP progress at the end of the school year and with standardized measures of the Behavior Assessment System for Children and Vineland Adaptive Behavior Scales (VABS). GAS scores were associated with parent and teacher ratings, including the VABS.

Keywords

Goal Attainment Scaling; Autism; IEP; outcome assessment; progress monitoring; Behavioral Assessment System for Children; Vineland Adaptive Behavior Scales

Comparative Study of Idiographic vs Nomothetic Outcome Measures for Students with ASD

Students with autism spectrum disorder (ASD) have difficulties in social and communication skills, as well as restricted and repetitive behaviors that interfere with early development and

Correspondence Lisa Ruble, Earl F. Smith Distinguished Professor in Special Education and Autism, Teachers College, TC745, Ball State University, Muncie, Indiana 47306, USA, Phone: 765-285-5702 laruble@bsu.edu.

Author note

Lisa Ruble, Department of Special Education, Ball State University

John McGrew, Department of Psychology, Indiana University-Purdue University

Brittany Dale, Department of Special Education, Ball State University

Madison Yee, Department of Educational, School, and Counseling Psychology, University of Kentucky

Compliance with Ethical Standards

Ethical approval: All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments.

Informed consent: Informed consent was obtained from all individual participants included in the study.

self-directed learning (American Psychological Association, 2013). Nearly two decades have passed since the National Research Council (2001) recommended ASD-specific content areas that should be included in early intervention educational programs to address the underlying impairments associated with ASD. Although research-supported practices vary and broadly target some of these areas of slow- or under-development, intervention goals also must be individualized to the personalized needs of the child (McGrew et al., 2016). The common practice of choosing goals from a bank or library of skills fails to ensure that the selected educational goal is targeted and personalized to account for the specific child's risk and protective factors, i.e., the parent- and teacher-identified concerns that most severely interfere with or aid growth and development (More & Hart Barnett, 2014).

Although limited, research on how well public-school programs produce and address best practice recommended goals personalized to the child suggests that the Individual Education Programs (IEP) of young children with autism are in need of improvement (Ruble et al., 2010b; Slade et al., 2018; White, 2014). For example, ASD-specific nonacademic skills such as social, communication, and self-directed behaviors have clear pivotal impact on other areas of development (Koegel & Koegel, 2006) yet are often missing or insufficiently addressed in IEPs (White, 2014). Because of these overlooked or unaddressed needs, parents of children with autism are more likely to be dissatisfied with special education services and, as a result, to pursue mediation and due process regarding their child's IEP (White, 2014). White (2014) analyzed complaint investigations initiated by parents and identified IEP content as a frequent concern (71%), including other concerns such as a lack of measurable goals and objectives, outcome assessment, and progress monitoring.

These difficulties with personalized goal creation and measurement are not isolated to special education programming but apply to student and educational outcome assessment more generally (Anagnostou et al., 2015; McConachie et al., 2015). In a comprehensive review of psychometrically sound and socially validated measures or approaches to assess intervention progress and outcomes for nonacademic goals for students with ASD, McConachie and colleagues (2015) reviewed more than 3,000 papers that included outcome measures for young children with ASD age six and younger, and noted both a paucity of measures available to assess educational or intervention outcomes; and, of those identified, limited to no evidence of reliability or sensitivity to track child progress over time or as a result of an intervention.

One possible reason for this poor state of affairs is that the assessment of nonacademic goals such as social communication and learning skills present unique challenges for special educators because of the complexity required for personalizing and individualizing goals for students with ASD (Etscheidt, 2003; Ruble et al., 2019). Unlike academic areas such as reading, writing, and math, where there are standardized, norm-referenced and curriculum-based assessments available to compare student performance against the performance of other students, the social communication and learning skills necessary for young children with ASD require personalized goal selection and intervention planning where outcome monitoring based on comparisons to a set of norms or scores from a population of students who have completed the same items is neither available nor realistic (McConachie et al., 2015). Further, the most common approach for evaluating student progress is based on a

standardized or group comparison framework, that is, nomothetic assessment. However, standardized assessments are problematic because they are less sensitive in observing changes in functioning and goal progress for individual students over time (McConachie et al., 2015) and fail to account for progress when students start at different levels of baseline performance, have different intervention plans, and have various targeted intervention outcomes (Ruble et al., 2019). Further, when goals and strategies are individualized, group assessment approaches are insensitive because they lack alignment with instruction or intervention plans. Although these concerns with individualized assessment have been acknowledged within special education over several decades (Carr, 1979; Shuster et al., 1984) the need for alternative outcome measures based on an idiographic approach has not yet been fully accepted or adequately addressed.

Standardized, norm-referenced measures are commonplace in special education and relied upon to make important educational placement decisions. School psychologists, occupational therapists, and speech language pathologists provide input on a child's educational needs through use of skills specific assessments. Given the frequent use of norm-referenced assessments, their familiarity among school professionals, and their generally sound psychometric properties (i.e., validity, reliability, etc.), it is not surprising that they have also been used to help identify progress toward intervention goals. Many of these tests, however, were neither designed nor standardized to serve as progress monitoring tools (McConachie et al., 2015) thus challenging the robust psychometrics generally presumed and accepted of these assessments. For instance, the Behavior Assessment System for Children (BASC-2; Reynolds & Kamphaus, 2004) was designed to be part of a multi-method, multi-informant evaluation to assist in differential diagnosis or education classification decisions and to help determine problem behaviors for treatment planning. It was not designed to monitor the progress of interventions and measure behavior change over time, and using it in that capacity incorrectly presumes utility of the assessment as being sensitive to measuring change. Furthermore, nomothetic approaches do not take into consideration that children with ASD will react to situations differently and typically pursue an individualized rather than a standardized educational program, weakening the sensitivity of group-based measures to detect progress (Ardion et al., 2008). The psychometric theory applied to these measures assumes group consistency of instructional behaviors across persons and situations causing them to fall short when measuring the effectiveness of interventions targeted to improve individualized, specialized aspects of ASD, such as social communication skills.

Compared to nomothetic approaches, idiographic approaches for individualized outcome measurement are an ideal alternative. The most common idiographic approach available is Goal Attainment Scaling (GAS) (Shuster et al., 1984). GAS was first described as a means of evaluating mental health treatment outcomes in 1968 by Kiresuk and Sherman and for assessing special education outcomes in 1979 by Carr. Using GAS, goals are benchmarked across five anchors from -2 (current level of performance) to $+2$ (most favorable outcome likely). The current level of performance represents the most unfavorable outcome, corresponding to no change or regression in level of performance. The midpoint, 0 level, represents an outcome equal to the expected level of performance at the end of the intervention or school year. The -1 level represents a level of performance that is

less than the expected outcome and a +1 represents a level of performance that is greater than the expected outcome. Since its inception, more than 1,000 articles have been written about GAS, including an increasing usage in assessing educational interventions (Shankar et al., 2020), such as the study of consultation outcomes (Hurwitz et al., 2015; Sladeczek et al., 2001), and ASD specific studies, including large quasi experimental designs (Odom et al., 2013), randomized controlled trials (Pfeiffer et al., 2011; Ruble et al., 2010a; Ruble et al., 2018; Ruble et al., 2013; Schaaf et al., 2014; Schlosser, 2004), classroom-based proof-of-concept studies (Schlosser et al., 2020), and single-subject designs (Charney et al., 2018; Oren & Ogletree, 2000).

GAS was the primary outcome measure applied in three RCTs of COMPASS, a parent-teacher consultation intervention for students with ASD. COMPASS is a manualized goal setting, intervention planning, and teacher coaching intervention (Ruble et al., 2012) and consists of a 3-hour parent-teacher consultation followed by four teaching coaching sessions dispersed equally throughout the remainder of the academic school year (< 10 hours total). The initial consultation is critical for obtaining a common understanding of the student's risk and protective factors at home and school from the parents' and teachers' perspectives. A COMPASS profile of student strengths and challenges associated with personal management, social skills, communication, learning skills, interfering behaviors, and sensory preferences and avoidances are reviewed (<https://compassforautism.org/build-a-compass-profile-2/>). A shared decision-making approach is then used to select goals and interventions to ensure high quality consensus driven goal selection and intervention planning. Based on recommendations from the National Research Council (2001), three major goal skill areas are targeted: communication skills, social skills, and learning skills. Each goal is then developed into a measurable skill. For this study, these goals make up the adaptive skills that were compared against various measurements (nomothetic and idiographic) and sources (parent and teacher) for data analysis.

Despite the increasing popularity of GAS and its use in RCTs such as COMPASS, questions remain about its construct and social validity (Shankar et al., 2020) (Ruble et al., 2012). Many of the concerns about its use within a group comparison design were addressed in a prior paper introducing Psychometrically Equivalence Tested Goal Attainment Scaling (PET-GAS, Ruble et al., 2012, described below in measures section). PET-GAS accounts for possible differences in outcome that could be accounted for by differences in the equivalence of the selected goals between the experimental and comparison groups.. Using features that could explain differences in performance between groups, PET-GAS ensures psychometric equivalence of GAS goals. The features we were interested in for educational-related GAS included level of difficulty, equidistance, and measurability. The difficulty of achieving a goal relative to baseline performance, the distance between goal benchmarks, and the measurability of the goal could account for differences in ratings and measurement error. That is when conducting RCTs using GAS, differences in outcomes between groups could be explained by these aspects, such that experimental group GAS ratings could be accounted for by easier to obtain goals that are less difficult, of shorter distance to the goal, and of poor measurability. Thus, we developed a GAS procedure to evaluate these features and renamed this PET-GAS, which is explained in further detail in the next section.

The current study focused on questions of construct and discriminant validity. Thus, the objective of this study was to determine the association between parent and teacher self-report ratings of child progress on IEP goals reflecting adaptive skills of social, communication, and learning skills at the end of the school year, and an objective assessment of the same goals from an independent observer using PET-GAS. Teacher and parent impressions of student progress are easily assessed and either conjointly or independently may provide an additional source of valid information about student's IEP goal progress. We also examined the PET-GAS change scores against nomothetic change scores assessing similar adaptive skills domains using the Behavioral Assessment System for Children-2 (BASC; Reynolds & Kamphaus, 2004) parent and teacher report and the Classroom Edition of the Vineland Adaptive Behavior Scales (VABS; Sparrow et al., 2005). One of the principles of good assessment is that triangulation using multiple sources often produces a more accurate picture (Ghrayeb et al., 2011). Thus, it is informative to understand how different sources relate to each other and to a criterion. Lastly, we wanted to examine areas of discordance, or discriminant validity and examined PET-GAS change with change scores for IQ, daily living skills, and motor skills.

Method

Design

This study is a secondary analysis of data aggregated from three randomized control trials of the Collaborative Model for Promoting Competence and Success (COMPASS) (Ruble et al., 2010a; 2013; 2018). In the three randomized control trials (RCT) of COMPASS, large experimental vs. control group differences in individualized student outcomes were obtained, with Cohen's *D* effect sizes ranging from 1.1 to 2.0 using PET-GAS as the primary outcome. In this secondary analysis, we compared goal attainment scores collected by raters blind to group assignment to parents and teachers' ratings of child goal attainment progress and nomothetic measures.

Participants

Data from 89 participant triads were collected. Each triad consisted of a student with ASD, his/her caregiver, and special education teacher. The children's ages ranged from 3.2 to 20 years old ($M = 8.7$; $SD = 5.4$). Eighty-five percent of the children were male. To participate in the study, each child had a formal diagnosis of ASD as confirmed by The Autism Diagnostic Observation Scale (ADOS Modules 1 and 2; Lord et al., 2000) and received special education services in a public school under the category of autism. Student cognitive ability scores ranged from 24 to 131 with a mean score of 56.4 ($SD = 25.8$), adaptive behavior scores ranged from 25 to 120 with a mean score of 63.62 ($SD = 14.6$). One caregiver completed all the questionnaires for each family. Caregivers identified as 73.0 % White, 10.1 % Black, 2.2% Asian, and 4.5% listed Other. About 55% of caregivers reported a total household income below \$50,000. Ninety-eight percent of the teachers were female and reported an average of 6.9 years ($SD = 6.8$) of teaching experience.

Research team raters.—In the next section we describe procedures for administration of our measures. The research team member participants who developed the PET-GAS and

conducted the interviews were primarily doctoral students in clinical or school psychology, a full time research assistant, and the first author. For the interviews and ratings conducted at the end, an independent rater was selected who was not part of the weekly research team meetings, was unaware of group assignment, and not involved in any of the intervention or research implementation activities.

Measures and Procedures

As noted above, psychometrically equivalence tested goal attainment scaling (PET-GAS; Ruble et al., 2012) was used to evaluate progress on the goals by an independent evaluator who was blind to students' group assignments and intervention plans. Similar to the original instructions for creating a GAS, PET-GAS uses a 5-point rating scale to rate student goal progress. But unlike the original GAS, the anchors were modified to be more consistent with educational goal setting and language: $-2 = \textit{student's present levels of performance}$, $-1 = \textit{progress}$, $0 = \textit{expected level of outcome by the end of the school year}$, $+1 = \textit{somewhat more than expected}$, $+2 = \textit{much more than expected}$. To ensure psychometric equivalence, we used a rubric to standardize the process for writing each PET-GAS goal (see Appendix). Once the goals were written, we also assessed level of difficulty of the skill, measurability of the skill, and distance between the goal benchmarks. Specifically, between group analysis of mean scores using t-tests were conducted for each of the three areas. If a significant difference was observed and non-equivalence was detected, goals were modified to be more comparable. In the rare situation in which group equivalence differences remained, the specific area of non-equivalence was entered as a covariate in any group comparison analyses. To assess reliability, two independent coders rated goal equivalence for 20% of the PET-GAS forms for each domain. Interrater agreement using the sample ICC for single measures was excellent; 1.0 for measurability, .92 for equidistance, and .93 for difficulty. The appendix provides the definitions and scoring instructions for each of the three areas.

PET-GAS pre- and post-treatment ratings were based on video demonstrations, work samples, and/or other relevant student data collected. To assess reliability, two independent raters coded 33% of the goals. Using the sample intraclass correlation coefficient (ICC) for single measures, interrater agreement was .82 for the social domain, .86 for the communication domain, and .91 for learning skills goals. PET-GAS includes a validation step to ensure goal equivalence across all students in measurability, equidistance, and difficulty.

The mean score for the three targeted instructional domains were used in data analysis. Change scores were obtained by subtracting scores from the start of the school year from final scores obtained at the end of the school year.

Parents and teachers assessed IEP goal progress with a questionnaire administered at the end of the school year. Parents and teachers were asked to rate student progress for the three goals reflecting a social skill, communication skill, and learning skill compared to the beginning of the school year using a 5-point scale, where 1 = Not at all, and 5 = A great deal. The mean score for the three areas were used in data analysis.

The Behavior Assessment System for Children Second Edition (BASC-2) is a multimethod, multidimensional system used to evaluate the behavior of children and young adults with different versions for three age groups (2–5, 6–11, and 12–21) (Reynolds & Kamphaus, 2002). The BASC-2 is considered multidimensional given its ability to assess both adaptive and maladaptive dimensions of behavior and personality (Reynolds & Kamphaus, 2004). The BASC-2 has 16 primary scales; however, for this analysis, only the Adaptive Skills composite scale was used, based on the parent and teacher report versions. Internal consistency analyses yielded coefficient alphas generally in the .90s for the composite scales, and reliabilities generally in the .80s for individual scales across the teacher and parent forms (Reynolds & Kamphaus, 2004). The T scores were used for analysis, and scores obtained from the parent and teacher at the start of the school year were subtracted from the scores obtained at the end of the year.

The Vineland Adaptive Behavior Scales-II (VABS) - Classroom Edition (Sparrow et al., 2005) is a standardized tool that allows for the measurement of personal and social skills commonly needed for everyday living. A 244-item classroom version assesses adaptive behavior for children between the ages of 3 and 12 years. This edition is administered in the form of a questionnaire completed by a teacher. Four domains are evaluated and include socialization, communication, daily living skills, and motor skills. The internal consistency coefficient alphas for the domains range from .80 to .95. The standard total score was used for analysis. Scores obtained at the start of the school year were subtracted from scores gathered at the end of the school year.

Intellectual ability was assessed with the Differential Abilities Scale (DAS; Elliott, 1990). The General Conceptual Ability (GCA) subscore was used to measure overall cognitive ability. The internal consistency of the GCA across ages ranges from .89 to .95 for the Preschool Level and from .95 to .96 for the School-age level. The test-retest reliability of the GCA for both age levels ranges from .89 to .95 (Elliott, 1990). DAS pre and post scores were available only from students who participated in our first two RCTs reflecting children between the ages of 3 and 9 years. Intellectual ability was not assessed at both timepoints for the high school youth who completed our third RCT. Thus the IQ change scores from the DAS with younger children were used for analysis.

Data Analysis

Because we were interested in the associations between PET-GAS change scores and parent/teacher report of change over the school year, change scores were calculated for measures that were obtained at the start and end of the school year. PET-GAS was collected at both timepoints for all three studies. The parent and teacher survey obtained at the end of the school year was collected for RCTs 2 and 3. The Vineland and DAS scores were collected at both timepoints for studies 1 and 2; lastly, the BASC parent report was obtained pre and post for all three studies; BASC teacher report for both timepoints was collected for studies 1 and 2. Partial correlations, controlling for group assignment, were utilized to evaluate associations between the idiographic (PET-GAS and parent/teacher ratings) and nomothetic (VABS, BASC-2, IQ) measures. Results were considered significant when the probability was less than .05, two-tailed.

Results

PET-GAS correlations.

We examined both convergent and discriminant validity. With respect to convergent validity, when the analyses were restricted to the idiographic measures, there were significant positive correlations between PET-GAS change scores and both parent ($r = .34, p = .010$) and teacher ($r = .54, p = .000$) end of the school year IEP progress ratings. When the analyses were expanded to examine correlations between idiographic and nomothetic measures, PET-GAS change scores correlated with three VABS subscales: VABS adaptive behavior composite change score ($r = .36, p = .006$); VABS communication change score ($r = .27, p = .038$); and VABS socialization change score ($r = .35, p = .006$). However, PET-GAS did not correlate with parent- or teacher-rated BASC Adaptive Skill change scores or the Daily Living and Motor Skills change scores of the VABS (see Table 1). Further IQ change scores did not correlate with any of the idiographic ratings.

Analysis of the correlation between each the variables by group assignment revealed a significant correlation for PET-GAS change ($r = .43, p < .001$) and parent progress ratings ($r = .33, p < .05$). None of the other correlations were significant.

Parent-teacher agreement.

There was a strong positive correlation between parent and teacher idiographic ratings of end of year progress ($r = 0.43, p < .01$). Idiographic and nomothetic parent ratings did not correlate with any of the teacher-rated nomothetic measures. However, idiographic teacher-rated end of the year progress correlated significantly with nomothetic teacher-rated VABS motor skills ($r = .59, p < .005$). There also was a nearly significant trend correlation between teacher ratings of progress and VABS adaptive behavior composite ($r = .38, p = .064$) and socialization scores ($r = .38, p = .064$).

Discussion

Students with ASD require instruction in domains associated with the diagnosis and underlying areas of weakness - social, communication, and learning / work behavior skills. Further, the National Research Council (2001) in their summary of empirically supported interventions for children with autism indicate that effective programs target these areas of developmental difference. A challenge for special educators is monitoring progress toward goals based on these nonacademic domains of instruction. In this regard, goal attainment scaling is an idiographic approach with lots of promise for meeting the need for a valid and reliable outcome assessment approach.

In their review of validation practices for GAS, however, Shankar et al. (2020) identified several gaps in validity evidence for GAS, including a paucity of research on the overall construct validity of GAS. The purpose of this study was to examine PET-GAS validity by testing associations between PET-GAS obtained from an independent observer and parallel ratings from multiple sources (i.e., parents and teachers), and using multiple methods, including standardized instruments assessing similar adaptive skills. The results indicated significant positive correlations among the idiographic measures, specifically, between PET-

GAS ratings and teacher and parent perceptions of progress over the school year. The highest correlations concerned those between independent observer rated PET-GAS and teacher ratings of IEP progress ($r = .54$) followed closely by parent ratings of IEP progress ($r = .34$).

Agreement between parent and teacher report of goal progress and independent PET-GAS ratings helps establish convergent validity of the PET-GAS approach. The results of the study also suggest that simple parent and teacher ratings are promising, valid indicators of goal progress.

When analyzing the associations between idiographic and nomothetic approaches to measuring goal progress, the results were mixed. For example, one standardized measure, the teacher rated VABS adaptive composite, communication, and social change scores correlated moderately with PET-GAS ratings, whereas, parent- and teacher-rated BASC change scores did not. Further, both nomothetic measures, i.e., the VABS and the BASC, failed to correlate with parent and teacher IEP progress ratings, although there was a trend ($p < .1$) for teacher-rated VABS composite change scores and VABS Socialization scores to correlate with teacher progress ratings.

The correlation between idiographic measures such as PET-GAS and nomothetic measures such as VABS, provides preliminary evidence of construct validity of goal attainment scaling as a measurement approach. This association between approaches increases when the content of the items and subscales of the nomothetic measure are more closely aligned with the specific IEP goals. For example, the VABS, in comparison to the BASC, may possess subscales with greater content overlap with the National Research Council (2001) specified ASD instructional domains, which guide the COMPASS IEP goal selection.

In contrast, the lack of correlation between the PET-GAS and the BASC-2 is consistent with previous literature that nomothetic scales are limited in their ability to measure idiographic progress, such as IEP goal attainment. Several characteristics of the BASC-2 may account for this failure. For instance, the response choices may not be sensitive to small changes in behaviors when three or four options are available. To illustrate this point, imagine a child who was initially rated as “sometimes” engaging in an adaptive behavior such as cleaning up after themselves. After intervention, they may engage in this activity more frequently, however, a teacher or parent may not believe it is enough to score the child as “often” performing the skill. In this case, change occurred because improvement was observed, but this change was not picked up through the nomothetic scale. Additionally, far fewer items related to adaptive functioning and skills that accurately address those specifically targeted by the IEP goals are likely to be included in the BASC-2 compared to a more comprehensive measure of adaptive functioning. Of those items, even fewer may be related to a student’s individualized goal further creating discordance between measurement strategy and the targeted behavior change. That is, the measurement content of the small number of items within a nomothetic subscale is unlikely to sensitively and precisely assess a carefully targeted and personalized IEP goal.

In addition to evidence of convergent validity, discussed above, construct validity also requires demonstration of discriminant validity, i.e., evidence that the scale fails to correlate with constructs not expected to be associated with the GAS. Accordingly, we examined correlations between PET-GAS and measures expected to be discordant to the PET-GAS change scores. As predicted, IQ change scores, VABS motor skills change scores, and VABS daily living skills change scores failed to correlate with PET-GAS.

In sum, the idiographic measure of PET-GAS appears to be a sensitive and valid approach for monitoring student goal progress. The significant correlations between group assignment and PET-GAS change and parent progress rating effect sizes provides additional psychometric evidence that idiographic measures are more sensitive approaches for evaluating treatment outcomes. Evidence for construct validity was primarily demonstrated through high levels of convergent validity as verified by strong associations with parallel idiographic approaches, i.e., parent and teacher ratings of goal progress, and modest levels of association with nomothetic approaches, i.e., teacher rated VABS scores assessing broad areas of adaptive skill, social, and communication. The stronger associations with idiographic approaches is not unexpected, given that IEP success is predicated on changes in specific and highly individualized goals. Additionally, because PET-GAS utilizes an individualistic method in monitoring goal progress, educators can track specific progress milestones for each child.

Another finding of potential importance was the utility of teachers and parents as a potential source for student progress ratings. Teacher ratings of student performance were robust indicators of progress, as they correlated strongly with two alternate idiographic measures of progress, i.e., observer ratings, parent ratings.

One limitation of this study, as noted previously, the version of the BASC included in this study was not designed to measure the progress of interventions. In response to this limitation of the BASC, the most recent version, the BASC-3 (Reynolds & Kamphaus, 2015), includes the BASC-3 Flex Monitor, which is designed to monitor changes in behavior over time. Publishers indicate parent, teacher, and self-report forms are available, and can measure the effectiveness of behavioral interventions. Further research on the PET-GAS should include comparisons to the BASC-3 Flex Monitor to further establish construct validity and utility in measuring change.

Acknowledgements

We wish to acknowledge Alexis Rodgers and Yue Yu for assistance with data coding.

Appendix

GAS Development and Coding Instructions

Instructions: A 3-point Likert Scale is applied for each of the three dimensions described. Each scale has its own anchors.

Level of Difficulty		
1	2	3
Not at all difficult	Somewhat difficult	Very difficult
1 – The skill is very close to what the child is already described as able to perform in the present levels of performance		
2 – The present levels of performance indicates that the child is able to perform the skill in limited ways compared to what is written in the objective (limited people, prompts, or places); if PLEP says child has difficulty doing it, score a “2”		
3 – The present levels of performance indicates that the child is unable to perform skill with anyone, anywhere, or with any prompts compared to what is written in the objective		
Measurability		
1	2	3
Not at all measurable	Somewhat measurable	Very measurable
1 – None or only one indicator (prompt level, criterion for success; observable skill) is listed for in each description		
2 - Two of the three indicators (prompt level, criterion for success; observable skill) are provided		
3 – Describes all three indicators (prompt level, criterion for success; observable skill)		
Equality		
1	2	3
Not at all equal	Somewhat equal	Clearly equal
1 – None or only 1 of the three descriptions (–1, +1, +2) are equilibrated appropriately in reference to the targeted objective (which is the zero level). Descriptions of prompts do not follow increasing independence from –2 to +2 and go from environmental/visual to verbal/physical and /or the skill frequency increases or reduces by less or more than 50% (do not include the present levels of performance –2 description) relative to the zero level.		
2- Two of the three descriptions (–1, +1, +2) are equilibrated appropriately in reference to the targeted objective (which is the zero level) accordingly from verbal/physical to visual/environmental prompts or the frequency of skill increases or reduces by at least 50% relative to the targeted objective (do not include the present levels of performance –2 description).		
3 – All of the three descriptions (–1, +1, +2) relative to the targeted objective are equilibrated and scaled appropriately: hierarchy of skills are reduced from verbal/physical to visual/environmental or frequency of skill increases or reduces by at least 50% for the all descriptions.		

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Anagnostou E, Jones N, Huerta M, Halladay AK, Wang P, Scahill L, Horrigan JP, Kasari C, Lord C, & Choi D (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism*, 19(5), 622–636. [PubMed: 25096930]
- Carr RA (1979). Goal attainment scaling as a useful tool for evaluating progress in special education. *Exceptional Children*, 46(2), 88–95. [PubMed: 510329]
- Charney L, Aimes J, Apgar L, Buday A, Calverley N, Cotta N, Kearney C, Redmond A, & Spires T (2018). Using single–subject design and goal attainment scaling to measure improvement in social participation in children with autism spectrum disorder. *American Journal of Occupational Therapy*, 72(4_Supplement_1), 7211505149p7211505141–7211505149p7211505141.
- Elliott CD (1990). The nature and structure of children’s abilities: Evidence from the Differential Ability Scales. *Journal of Psychoeducational Assessment*, 8(3), 376–390.
- Etscheidt S (2003). An analysis of legal hearings and cases related to individualized education programs for children with autism. *Research and Practice for Persons with Severe Disabilities*, 28(2), 51–69.

- Ghrayeb O, Damodaran P, & Vohra P (2011). Art of triangulation: An effective assessment validation strategy. *Global Journal of Engineering Education*, 13(3), 96–101.
- Hurwitz JT, Kratochwill TR, & Serlin RC (2015). Size and consistency of problem-solving consultation outcomes: An empirical analysis. *Journal of School Psychology*, 53(2), 161–178. 10.1016/j.jsp.2015.01.001 [PubMed: 25746825]
- Kiresuk TJ, & Sherman RE (1968). Goal attainment scaling: A general method for evaluating community mental health programs. *Community Mental Health Journal*(4), 443–453. [PubMed: 24185570]
- Koegel RL, & Koegel LK (2006). Pivotal response treatments for autism: Communication, social & academic development. Paul H. Brookes. <http://www.loc.gov/catdir/toc/fy0608/2005023288.html>
- Lord C, Risi S, Lambrecht L, Cook EH Jr., Leventhal BL, DiLavore PC, Pickles A, & Rutter M (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223. [PubMed: 11055457]
- McConachie H, Parr JR, Glod M, Hanratty J, Livingstone N, Oono IP, Robalino S, Baird G, Beresford B, & Charman T (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder.
- More CM, & Hart Barnett JE (2014). Developing individualized IEP goals in the age of technology: Quality challenges and solutions. *Preventing School Failure: Alternative Education for Children and Youth*, 58(2), 103–109.
- National Research Council. (2001). *Educating children with autism* National Academy Press. <http://www.nap.edu/books/0309072697/html/>
- Odom S, Cox A, & Brock M (2013). Implementation science, professional development, and autism spectrum disorders. *Exceptional Children*, 79(2), 233–251.
- Oren T, & Ogletree BT (2000). Program evaluation in classrooms for students with autism: Student outcomes and program processes. *Focus on Autism and Other Developmental Disabilities*, 15(3), 170–175.
- Pfeiffer BA, 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.
- Reynolds CR, & Kamphaus RW (2002). *The clinician's guide to the Behavior Assessment System for Children (BASC)*. Guilford Press.
- Reynolds C, & Kamphaus R (2004). *Behavior Assessment System for Children*. Circle Pines, MN: AGS Publishing.
- Reynolds CR, Kamphaus RW (2015). *Behavior assessment for children: Third edition. (BASC-3)*.
- Schaaf RC, Benevides T, Mailloux Z, Faller P, Hunt J, Van Hooydonk E, ... & 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. [PubMed: 24214165]
- Schlosser RW, Goal attainment scaling as a clinical measurement technique in communication disorders: A critical review. *Journal of Communication Disorders*, 37, 217–239. [PubMed: 15063144]
- Schlosser RW, Shane HC, Allen A, Benz A, Cullen J, Chiesa L, Miori-Dinneen L, Koul R, & Pasupathy R (2020). Coaching a school team to implement the Visual Immersion System™ in a classroom for children with Autism Spectrum Disorder: A mixed methods proof-of-concept study. *Advances in Neurodevelopmental Disorders*, 4(4), 447–470.
- Shankar S, Marshall SK, & Zumbo BD (2020). A systematic review of validation practices for the goal attainment scaling measure. *Journal of Psychoeducational Assessment*, 38(2), 236–255. 10.1177/0734282919840948
- Shuster SK, Fitzgerald N, Shelton G, Barber P, & Desch S (1984). Goal attainment scaling with moderately and severely handicapped preschool children. *Journal of Early Intervention*, 8(1), 26–37. 10.1177/105381518400800104
- Slade N, Eisenhower A, Carter AS, & Blacher J (2018). Satisfaction with individualized education programs among parents of young children with ASD. *Exceptional Children*, 84(3), 242–260.

- Sladeczek IE, Elliott SN, Kratochwill TR, Robertson Mjaanes S, & Stoiber KC (2001). Application of goal attainment scaling to a conjoint behavioral consultation case. *Journal of Educational and Psychological Consultation*, 12(1), 45–58.
- Sparrow SS, Balla DA, & Cicchetti DV (2005). *Vineland adaptive behavior scales: Survey forms manual*. AGS Publ.
- White SE (2014). Special education complaints filed by parents of students with autism spectrum disorders in the midwestern United States. *Focus on Autism and Other Developmental Disabilities*, 29(2), 80–87.

Table 1.

Partial Correlations of PET-GAS and Parent and Teacher Ratings of Idiographic and Nomothetic Change Scores and Effect Size (ES) Correlations

	1	2	3	ES
Idiographic Measures				
1. PET-GAS Change	-			.43***
2. Parent Progress Rating	.34 [*] /56	-		.33 [*]
3. Teacher Progress Rating	.54***/44	.43 ^{**} /39	-	.14
Nomothetic Measures				
4. Teacher VABS Change	.36 ^{**} /54	-.02/31	.38 [^] /22	.11
5. Teacher VAB Comm Change	.27 [*] /54	.11/33	.14/23	-.11
6. Teacher VAB Social Change	.35 ^{**} /58	.19/35	.38 [^] /22	.19
7. Teaching VAB Daily Living Change	.21/57	.09/34	.29/23	.00
8. Teacher VAB Motor Change	.26/34	.07/23	.59 [*] /15	-.16
9. Parent BASC Change	-.05/75	.03/52	-.17/38	-.05
10. Teacher BASC Change	.03/80	.01/56	.03/44	-.10
Discriminant Validity				
11. IQ Change	-.18/62	.11/37	.12/24	.07

Note:

*
p< .05

**
p<.01

[^]
p<.1; two-tailed test; VABS is based on the Classroom Edition; BASC is based on Adaptive Composite; Change scores were calculated by subtracting Time 1 ratings collected at the start of the school year from Time 2 ratings collected at the end of the school year. The lower diagonal represents the df. ES calculated by correlating variables against group assignment (COMPASS vs Control).