Goal Attainment Scaling: Outcome Measure of Consultation and IEP Progress

Lisa Ruble
University of Kentucky

John McGrew
Indiana University-Purdue University

NASP 2011

Objectives/Learning Outcomes:

1. The use of GAS as a tool for:
   - Monitoring individual and group student progress and evaluating IEP outcomes
     - Teacher consultation
   - Conducting educational outcome research

2. How to take IEP objectives or other educational objectives and translate them into a GAS template.
Objectives/Learning Outcomes:

3. Understanding the assumptions of GAS and the strategies developed to test these assumptions GAS
4. A case study of GAS as an outcome measure applied to a randomized controlled study of teacher consultation

Progress monitoring and accountability

- IDEA and NCLB
  - Curriculum-based measurement
  - Classroom and large scale assessments

BUT......

How do we account for pivotal skills that impact all learning and yet are not well accounted for by standards-based assessment and accountability systems.
Challenge

- How do we monitor effectiveness of interventions and educational programs when
  - Skills represent
    - Social
    - Communication
    - Independent skills
  - The learning objectives (outcomes) are personalized for each student
  - Students start at different levels
  - The intervention varies

Goal Attainment Scaling

- GAS as a measurement system that can be applied as an objective approach for progress and outcome monitoring
  - Individual
  - Aggregated
- Sensitivity to the intervention being applied
So what about GAS?

- Allows for measurement of outcomes that are customized to the context/individual.
  - Been applied to district, school, classroom, and child level outcome assessment
  - Can be used with students who have different intervention outcomes and plans.
- Produces a Goal Attainment Score (GAS) which allows you to track progress and compare progress between groups.
- Allows you to weight goals according to importance.

Overview of Goal Attainment Scaling

- Developed by Kiresuk and Sherman (1968)
  - For mental health practitioners initially
  - Used by an array of disciplines today
- Used as the main outcome in studies on consultation effectiveness and is considered to be a standard (e.g., Sladeczek, et. al., 2001).
- Blends well with IEP goals (Oren & Ogletree, 2000).
Steps to Create GAS: Standard description

1. Identify expected goals/outcomes
2. Weight the goals according to priority (more severe, more importance = higher weighting)
3. Identify continuum of benchmarks
4. Determine baseline performance
5. Implement intervention
6. Monitor progress
7. Evaluate final goal attainment

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much less than expected outcome</td>
<td>-2</td>
</tr>
<tr>
<td>Less than expected outcome</td>
<td>-1</td>
</tr>
<tr>
<td>Expected outcome</td>
<td>0</td>
</tr>
<tr>
<td>More than expected outcome</td>
<td>+1</td>
</tr>
<tr>
<td>Much more than expected outcome</td>
<td>+2</td>
</tr>
</tbody>
</table>
Expected Outcomes

- Start by identifying the most likely outcome.
- Might be based on annual outcomes or broken down for every 6 weeks
- The middle level is the most probable and successful level of goal attainment
- Above this level is even greater response
- Below this level is less successful response
More about goal writing:

- The outcomes should be:
  - measurable and specific
  - not too easily accomplished or too difficult to obtain
- Two independent observers should be able to agree on whether it has been obtained
  - Adequate interobserver reliability

Scoring Goal Attainment Scales

- If you are working with individuals:
  - Each time you measure goal attainment mark the box which indicates the individual’s current status.
  - To obtain a score, add up the scores for each goal.
  - Scores on each goal range from -2 to +2 with 0 indicating treatment success.
  - The total across all goals represents the goal attainment score.
Scoring

- All scores are standardized and converted to T-scores (i.e., $M = 50; SD = 10$) using the Kiresuk-Sherman formula (Kiresuk, et. al., 1994).

If you are comparing or collecting scores from a number of individuals with different numbers of goals:

- Convert the GAS into a standard score.
- Refer to the GAS conversion table.
- Find the row which indicates an individual's goal attainment score and find the column which represents the number of goals.
- The standardized GAS score is where the row and column intersect.
Assumptions of GAS

- GAS is a continuous measure
- Comparability of GAS scores across individuals
  - Equal interval between each scaled description
  - Measurability or objectivity of the targeted skill
  - Degree of difficulty in obtaining the skill due to different levels of starting abilities
- Argued as a nonstandard measure (Mackay, 1996)

- Reliability of progress
  - Subjectivity of ratings (parent & teacher report)
Suggestions

- Conduct training on writing GAS
  - degree of difficulty and equality of intervals
- Operationalize definitions of outcome criteria
  - measurability
- Collect subjective and objective data in the determination of attainment levels
- Utilize an independent observer to code GAS
- Ensure adequate treatment integrity


Case Example

- The limitations of GAS and strategies to test the assumptions of GAS as a standard approach are presented within a randomized controlled study of teacher consultation for students with autism.
Our adaptation for use of consultation outcome research on IEP goal attainment

- A 5-point response scale is used:
  - -2 (child’s present level of performance = worst outcome)
  - -1 (progress)
  - 0 (expected level of outcome)
  - +1 (somewhat more than expected)
  - +2 (much more than expected)

- For each goal, specific behavioral descriptors were developed delineating degrees of progress toward the goal.
- Completed before intervention began.

### Goal Attainment Scale

<table>
<thead>
<tr>
<th>-2* Present level of performance</th>
<th>-1 Progress</th>
<th>0 Expected level of outcome (goal)</th>
<th>+1 Somewhat more than expected</th>
<th>+2 Much more than expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Present levels is used because the children are not expected to regress as recommended by Schlosser, 2004*
Creating a GAS form

- A protocol was developed to ensure that the descriptions were measurable and of equal interval.
  - Completed after the fact, as a means for ensuring forms were similar between groups.

Addressing Untested Assumptions of GAS

1. Comparability:
   Developed 3 dimensions coded by an independent rater:
   - Measurability
   - Level of Difficulty
   - Degree of equidistance between intervals
2. Objectivity/Reliability:
   - GAS coding was based on direct observation
   - Reliability of scores from live coding vs taped coding
     - Reliability of scores between teacher made tapes and live coding by researcher
     - Do teachers provide the “best scenario” tapes

What Problem are We Trying to Address Using GAS?

- Autism is described as a “national health emergency” (Interagency Autism Coordinating Committee, 2009)

- Annual costs are $35-90 billion (Jarbrink & Knapp 2001; Montes & Halterman, 2008)

- Total lifetime costs for individuals of 3.2 million (Ganz, 2006)
The Problem

- Compared to other children with special health care needs, children with ASD unmet needs for
  - specific health care services,
  - family support services,
  - delayed or foregone care,
  - difficulty receiving referrals, and
  - care that is not family centered
- African American and Latino children
  - receive later diagnosis,
  - have greater symptom severity, and
  - receive significantly lower numbers of services (Liptak, 2008; Mandel, et al., 2006, 2009; Morrier, et al., 2008)
- Children living in rural areas are underserved (Chen et al., 2008; Farmer, et al., 2005)

About 1/100
Centers for Disease Control
Solutions?

- Where are children getting services?
- What are mandated services?

Educational system is the only public funded and mandated agency to provide services

But...

- National shortage of special educators
- Educators lack knowledge of evidence based practices (Hess et al., 2008; Stahmer et al., 2005)
  - Services are often insufficient, lack specificity and intensity
  - Geographic disadvantages
- Little research available from community-based settings on effectiveness
The Question

- How can we start to reduce the research-to-practice gap by addressing the educational outcomes of children with autism?

Consultation

- **Teacher consultation is effective** (Busse et al., 1995; Medway & Updyke, 1985; Sheridan et al., 1996)

- **Models of consultation**
  - Behavioral (Noell et al., 2005)
  - Conjoint behavioral (Freer & Watson, 1999; McDougal, Nastasi, & Chafouleas, 2005; Sheridan & Steck, 1995; Sheridan, Clarke, Knoche, & Edwards, 2006; Sheridan; Eagle, Cowan, & Mickelson, 2001; Sladeczek, et. a., 2001; Wilkinson, 2005)
  - Systems consultation (Denton et al., 2003)
Types of problems

- Student achievement (Givens Ogle et al., 1991)

- Student disruptive behavior (Denton et al., 2003; McDougal et al., 2005; Ray et al., 1999; Sheridan et al., 2001; Sladeczek et al., 2001; Wilkinson, 2005)

- Teacher behavior (Cossairt, Vance Hall, & Hopkins, 1973; Meyers, Freidman, & Gaughan, 1975; Noell et al., 2005; Sparks, 1988; White & Fine, 1976)

- Parent-teacher relationships (Sheridan et al., 2006).

Primary outcome measure

- **Goal attainment scaling**
  - Start at different levels, different outcomes, different interventions

- **Address core symptoms of autism**
  - Impaired communication
  - Impaired social interaction
Randomized Single Blind Controlled
(Ruble, Dalrymple, & McGrew, 2010)

Time 1
Baseline Evals
Randomized (n = 35)

Control (n = 17)

COMPASS + Coaching (n = 18)

Time 2
Goal Attainment Scaling
*Unaware of Group Assignment

Group Comparison

- Comparison Group
  - Services as usual
  - Final evaluation

- Intervention Group
  - 3 hour COMPASS consultation (parent and teacher)
    - 3 IEP objectives
      - Specific to needs of students with autism
      - Measurable
      - Teaching plans
  - 4 teacher coaching sessions (1.5 hour every 4-6 weeks)
  - Final evaluation
## Between Group Comparisons

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control</th>
<th>Experimental</th>
<th>t(df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Child</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>5.98 (1.5)</td>
<td>6.18 (1.9)</td>
<td>-.34(33)</td>
<td>.74</td>
</tr>
<tr>
<td>Childhood Autism Rating Scale</td>
<td>41.43 (8.2)</td>
<td>36.38 (9.9)</td>
<td>1.55(30)</td>
<td>.13</td>
</tr>
<tr>
<td>Differential Abilities Scale</td>
<td>39.47 (18.4)</td>
<td>53.78 (27.1)</td>
<td>-1.81(33)</td>
<td>.08</td>
</tr>
<tr>
<td>Oral and Written Language Scales</td>
<td>41.13 (19.0)</td>
<td>51.56 (17.2)</td>
<td>-1.68(32)</td>
<td>.10</td>
</tr>
<tr>
<td>Adaptive Behavior Scales</td>
<td>62.29 (9.2)</td>
<td>64.88 (16.7)</td>
<td>-.56(32)</td>
<td>.58</td>
</tr>
<tr>
<td>BASC (externalizing composite)</td>
<td>59.53 (8.5)</td>
<td>59.83 (7.0)</td>
<td>-.11(31)</td>
<td>.91</td>
</tr>
<tr>
<td><strong>Teacher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Number of Children Taught</td>
<td>8.85 (11.5)</td>
<td>4.56 (6.1)</td>
<td>1.29(27)</td>
<td>.21</td>
</tr>
<tr>
<td>Total Years Autism</td>
<td>8.27 (8.3)</td>
<td>5.34 (5.5)</td>
<td>1.16(29)</td>
<td>.25</td>
</tr>
</tbody>
</table>

## Primary Hypothesis

Children whose teacher’s and parent’s participated in the consultation intervention would demonstrate better goal attainment outcomes compared to children who received their usual education program.
1. When presented with a task menu, Anthony will start and complete three 2-3 minute tasks each day without aggression with one adult verbal cue (e.g., time to work) and gestural/picture cues across two weeks.

2. During structured play, Anthony will imitate adult play activities for five actions (actions with objects) with at least three different preferred objects (dinosaurs, animals, doll) each day across two weeks.

3. Anthony will make 10 different requests per day independently (go home, eat, help, more, finished, various objects/activities) or as a response to a question (“what do you want?”) using sign, pictures, or verbalization.
### Progress Descriptions

<table>
<thead>
<tr>
<th>Dimension</th>
<th>GAS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of targeted skill</td>
<td>Lowest, Highest</td>
</tr>
<tr>
<td>Frequency of prompting</td>
<td>Highest, Lowest</td>
</tr>
<tr>
<td>Form of prompting*</td>
<td>Physical, Independent visual supports</td>
</tr>
<tr>
<td>Context†</td>
<td>Structured, Unstructured</td>
</tr>
<tr>
<td>Person</td>
<td>An adult, Many adults</td>
</tr>
</tbody>
</table>

*Note: "Physical" refers to social, communication, housing skill, adaptive behavior, motor/ sensory or academic goals.
†Note: "Structured" refers to at least two words"
### Goal Attainment Scale (GAS) Form for Anthony

<table>
<thead>
<tr>
<th></th>
<th>Present level of performance</th>
<th>Progress</th>
<th>Expected level of outcome (GAS)</th>
<th>+1</th>
<th>+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aggression when given a task he does not want to do. Is difficult to motivate. Does not have a non-appropriate way to communicate needs or to negotiate.</td>
<td>When presented with a task, Anthony will start and complete only one of two 5-minute tasks. Will not use aggression at any time.</td>
<td>When presented with a task, Anthony will ven start and complete three 5-minute tasks. Will not use aggression at any time.</td>
<td>When presented with a task, Anthony will complete three 5-minute tasks. Will not use aggression at any time.</td>
<td>When presented with a task, Anthony will start and complete more than three 5-minute tasks. Will not use aggression at any time.</td>
</tr>
<tr>
<td>2</td>
<td>Has difficulty manipulating visual stimuli. May use aggression as a way to request. Be lim on adult prompting to make requests.</td>
<td>Play activities for five minutes with at least three different preferred object (decoration, animals, doll...) each day across two weeks.</td>
<td>Play activities for five minutes with at least three different preferred object (decoration, animals, doll...) each day across two weeks.</td>
<td>Play activities for five minutes with at least three different preferred object (decoration, animals, doll...) each day across two weeks.</td>
<td>Play activities for five minutes with at least three different preferred object (decoration, animals, doll...) each day across two weeks.</td>
</tr>
</tbody>
</table>

### Interobserver Reliability of GAS:

- An observer scored the GAS form on 20% of the sample and results were compared with the primary observer’s scores (both direct observation coding).
- Weighted kappa coefficient of .65, indicating moderate to good agreement.
Fidelity

96% (teacher)
83% (parent)

Teacher adherence: (5 point scale)
1 “0% implementation” to 5 “100% implementation”

<table>
<thead>
<tr>
<th>Coaching Session</th>
<th>Fidelity Score Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0 (2.0)</td>
</tr>
<tr>
<td>2</td>
<td>3.5 (1.3)</td>
</tr>
<tr>
<td>3</td>
<td>4.0 (1.2)</td>
</tr>
<tr>
<td>4</td>
<td>4.1 (1.2)</td>
</tr>
</tbody>
</table>
Interventions Outside School

- No significant difference between groups for the number (t (27) = -.83, n.s.) or hours (t (33) = -1.00, n.s.) of services received outside of school.

Results

- Do children whose teachers receive COMPASS have better IEP goal attainment outcomes?
The level of difficulty in the GAS forms was lower for the consultation group vs comparison so they can’t be compared.

The intervals between the scales were not equivalent so group differences can’t be compared.

The measurability of the objectives was poorer for the comparison group so they can’t be compared.

*Raw GAS change scores: t(27) = -2.6, p = .02, d = 1.0
Group Comparisons

- Operationalized measurability, difficulty, equidistance
- Two raters coded GAS forms separately until 80% agreement
- Rater unaware of group assignment did final coding
- All coding was based on observation instead of parent/teacher report

Definitions: Measurability

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>None or only one indicator (prompt level, criterion for success; observable skill) is listed</td>
</tr>
<tr>
<td>2</td>
<td>Two of the three indicators (prompt level, criterion for success; observable skill) are provided</td>
</tr>
<tr>
<td>3</td>
<td>Describes all three indicators (prompt level, criterion for success; observable skill)</td>
</tr>
</tbody>
</table>
### Definitions: Difficulty

<table>
<thead>
<tr>
<th>Description</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skill is very close to what the child is already described as able to perform in the present levels of performance.</td>
</tr>
<tr>
<td>2</td>
<td>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”.</td>
</tr>
<tr>
<td>3</td>
<td>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.</td>
</tr>
</tbody>
</table>

### Definitions: Equidistance

<table>
<thead>
<tr>
<th>Description</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>None or only 1 of the three descriptions are equilibrated appropriately in reference to the targeted objective (which is zero). [Prompts described as going from environmental to verbal, and/or the skill frequency increases or reduces by less or more than 50% (do not include the present levels of performance -2 description)].</td>
</tr>
<tr>
<td>2</td>
<td>Two of the three descriptions are equilibrated appropriately in reference to the targeted objective (which is zero). [Descriptions are scaled accordingly from verbal to 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)].</td>
</tr>
<tr>
<td>3</td>
<td>All of the three descriptions relative to the targeted objective are equilibrated and scaled appropriately. [Hierarchy of skills are dropped from verbal to physical to visual/environmental or frequency of skill increases or reduces by at least 50% for the majority].</td>
</tr>
</tbody>
</table>
Intercorrelations between dimensions

<table>
<thead>
<tr>
<th></th>
<th>Measurability</th>
<th>Difficulty</th>
<th>Equidistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurability</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>.06</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Equidistance</td>
<td>-.06</td>
<td>.03</td>
<td>--</td>
</tr>
</tbody>
</table>

T-test Analysis of GAS Scores*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Comparison (n=34)</th>
<th>Consultation (n=48)</th>
<th>t-test</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurability</td>
<td>2.7 (.52)</td>
<td>2.5 (.62)</td>
<td>-1.5, p = .23</td>
<td>.35</td>
</tr>
<tr>
<td>Difficulty</td>
<td>2.3 (.60)</td>
<td>2.1 (.66)</td>
<td>1.9, p = .06</td>
<td>.32</td>
</tr>
<tr>
<td>Equidistance</td>
<td>2.9 (.34)</td>
<td>2.8 (.42)</td>
<td>.29, p = .78</td>
<td>.26</td>
</tr>
</tbody>
</table>

Original results the same, after controlling for level of difficulty ANCOVA, F (1,70) = 21.0, p .000, d = 1.0

Note: Comparison group GAS forms were masked to make them measurable

*Based on GAS forms at baseline for the comparison group and after consultation for the experimental group
Current Study
American Recovery & Reinvestment Act NIH Challenge Grant 5RC1MH89760

- Same experimental design and participant characteristics, but...
- Three groups
  1. COMPASS + FF Coaching (3 hours + 4 sessions, about 1 - 1.5 hours each)
  2. COMPASS plus WEB-based coaching
  3. Online teacher training (instead of SAU)
- All teachers did own taping for coaching sessions (we still did direct observation for final determination of progress at end of year)

Additional GAS issues when videotaping is used...

- Parent and teacher report is subjective
  - When teachers make and provide their own videotapes of instruction, do they take the “best” example?
  - We collected extra data point (live) that corresponded to a teacher coaching session

- Is coding based on live observation and from taped observation reliable?
  - Each has own advantages/disadvantages
  - Live allows more information of the context
  - Taped allows for observation to be played back
Opportunity to explore additional GAS issues when using videotape analysis

- Scores from teacher supplied taped instruction (n=11) was reliable with researcher supplied taped instruction
  > ICC = .65

- Scores based on coding done from live observation and coding from the same observation but taped were reliable
  > ICC = .66

But what about the three dimensions of GAS?*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Comparison (n=21)</th>
<th>Consultation (n=31)</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurability</td>
<td>2.9 (.30)</td>
<td>2.6 (.49)</td>
<td>2.4, p = .02</td>
</tr>
<tr>
<td>Difficulty</td>
<td>2.2 (.44)</td>
<td>2.4 (.56)</td>
<td>-1.3, p = .20</td>
</tr>
<tr>
<td>Equidistance</td>
<td>2.8 (.40)</td>
<td>2.7 (.44)</td>
<td>.56, p = .58</td>
</tr>
</tbody>
</table>

*Based on GAS forms at baseline for the comparison group and after consultation for the experimental group; coded at end of the school year during final evaluation by independent observer
Bottom Line

- Do teacher present the base case scenarios when they provide the taped observation?
  - No – time is of the essence
    - We are only using researcher supplied tapes for GAS scores for research purposes

- Are scores based on tapes reliable with scores based on live coding? Yes

- Is it necessary to monitor the dimensions of measurability, level of difficulty, and equidistance? Yes and early on

Acknowledgements

- Teachers, Indiana & Kentucky
- Families and Children
- Nancy Dalrymple, Co-investigator
- Michael Toland, UK
- Research team, UK
  - RAs: Ryan Johnson & Lauren Feltner
  - GRAs: Rachel Aiello, Jessie Birdwhistell, Jennifer Hoffman
  - Co-investigators: Jennifer Grisham Brown & Lee Ann Jung

- Research was supported by Grants No. R34MH073071 and 1RC1MH089760 from the National Institute of Mental Health.