ASSESSMENT OF "VALUE ADDED" BY SCHOOL PSYCHOLOGISTS
Louisiana Department of Education & Louisiana School Psychological Association
Task Force Recommendations

June 2010

Debra Duhe, *Supervisor of School Psychological Services, Louisiana Department of Education*

Cindy D'atrio, *President, Louisiana School Psychological Association*

Kevin Jones, *President-Elect, Louisiana School Psychological Association*

George Hebert, *LAS*PIC Coordinator*

Yael Banai, *LSPA Executive Council*

Connie Cassels, *LSPA Executive Council*

Jim Cowie, *LSPA Executive Council*

Michael Welch, *LSPA Executive Council*

OVERVIEW:

In May 2010, Louisiana adopted House Bill 1033 requiring annual evaluations of *all* teachers and administrators, and requiring that by 2012-13, 50% of these evaluations be based on evidence of growth in student achievement or "value added." The Louisiana School Psychological Association (LSPA) and Louisiana Department of Education (LDOE) Task Force was charged with the development of a formal plan for evaluating "value added" by certified school psychologists currently employed by public schools. There is currently no state department in the U.S., to our knowledge, that evaluates school psychologists using direct measures of student outcomes, such as changes in test scores. Thus, the recommendations of this task force were based on various approximations to this model, including (a) the standards set by the National Association of School Psychologists (NASP) for the evaluation of student outcomes, (b) standards published by the nation's only state-supported internship program in Ohio, and (c) a literature review of "best practices" in professional accountability. Based on this review, the task force proposal includes four recommendations that will be submitted for further review by all training institutions in Louisiana as well as all members of LSPA:

Recommendation 1. *Every school psychologist certified in Louisiana will submit three Value Added Portfolios (VAP) to the Louisiana Department of Education each year.*

A Value-Added Portfolio (VAP) is a case study that includes a summary of child characteristics, problem solving steps, a graph representing baseline and treatment data, and statistical evidence of improved student outcomes. This requirement is aligned with NASP standards (for those practitioners submitting evidence for national certification), and is the central element of existing accountability models in other states. The criterion of three cases is based on the number of single case design studies recommended by the American Psychological Association (APA) for establishing whether a treatment is "evidence based," and thus is likely to represent an adequate sample for accountability purposes. Although fewer cases may also be sufficient, a larger sample is unrealistic due to the nature of school psychological services. In practice, school psychologists must demonstrate outcomes *beyond the effects of general or special education*; if traditional methods were sufficient, the child would not be referred for additional supports. Thus, an individual child who receives *individualized* psychological services may require the same amount of resources and time as a group of children who receive *group* instruction.

The VAP will provide evidence of value added by demonstrating that intervention supports coordinated by the school psychologist led to improved student outcomes in a targeted domain. Given the wide range of potential educational and mental health outcomes, it is important that a value added metric be sufficiently flexible to accommodate a host of concerns, including motivation problems, academic skills, low versus high frequency behaviors, and acute/crisis versus long-term programs. For this reason, the value added metric featured in this proposal will allow for both quantitative and subjective evaluation.

Finally, the level or tier of service delivery is expanded to include "cases" that do not necessarily require direct contact with the student. One of the primary challenges for adopting a standardized "value-added" program for school psychologists is the diversity of service delivery roles and placement demands confronted by practitioners in Louisiana. Many practitioners hold supervisory and administrative positions that allow no direct contact with children and families, while others serve sparsely populated districts or assume specialties that restrict intensive and ongoing services. Although this issue will impact value-added models *for all disciplines*, the task force has anticipated special circumstances, and defined broadly the role of the school psychologist in the coordination of cases. A VAP submitted by a

practitioner may, for example, include a case in which the school psychologist's role was the supervision of services delivered to the student, or a case in which the practitioner provided indirect or team-based consultation to teachers or parents, who themselves serve as the primary change agent.

To qualify for submission, the case must be one in which the applicant is a major contributor to the coordination of psychological services. This would apply to cases in which the applicant is the provider or supervisor of intervention activities, including the design, implementation, and evaluation of the intervention. All of the following examples would be acceptable:

1. The school psychologist administers treatment directly to a child or a group of children.

2. The school psychologist provides training and supervises a treatment delivered by a classroom peer, a cross-age peer, an aide, or a practicum student.

3. The school psychologist provides dyadic consultation to a parent, teacher, or administrator, who delivers treatment directly to a child or group of children.

The elements of a VAP include a title page and a case study summary.

Page 1: **Title Page**– application information using an approved DOE form that includes a completed checklist of demographic/background information relevant to case.

Pages 2 – 5: **Case Study** – a case report that summarizes procedures and outcomes for a case that falls within any one of six service delivery domains (see Appendix A for definitions and examples of each domain):

	Cognitive/Academic Target	**Social/Affective Target**
Level 1: Universal	CAT1: A universal program that targets a cognitive/academic outcome.	SAT1: A universal program that targets a social/affective outcome.
Level 2: Selective-Group	CAT2: A selective-group remediation program that targets a cognitive/academic outcome.	SAT2: A selective-group remediation program that targets a social/affective outcome.
Level 3: Individualized	CAT3: An individualized intervention that targets a cognitive/academic outcome.	SAT3: An individualized intervention that targets a social/affective outcome.

Case studies include the following content in a narrative report that features the following sections:

- A **Referral Concern** section that introduces child characteristics, including age, grade, and reasons for referral
- A **Problem Identification** section that details the primary outcome measure used to monitor and evaluate progress
- A **Problem Analysis and Intervention Design** section that describes how treatment was selected, the procedural steps, and how treatment fidelity was ensured

- A **Results** section that displays a graph or figure, summarizes visual inspection, and includes the calculation of summary statistics that establish improved student outcomes

Each VAP shall be submitted electronically, as a single file document, to the Louisiana Department of Education.

Recommendation 2. The state department will contract with university-based teams to evaluate Value Added Portfolios (VAP).

Value added metrics used for school psychologists must be generated and evaluated on a case by case basis. This proposal will result in the submission of three VAPs per year from each of Louisiana's 300 school psychologists. Statewide training initiatives and the evaluation of 900 VAPs per year will require substantial expertise and resources. The evaluation component alone will require approximately 30 minutes each, using a standardized scoring rubric adopted from NASP. There are ten criteria for successful portfolios, and all must be met:

Complete Problem Solving Steps: The case summary provides a sufficient description of all of the following:
1. Child Characteristics (age, grade, demographics)
2. Problem Identification (procedures used to monitor progress, how often, how it was scored)
3. Problem Analysis and Intervention Design (treatment selection, procedural steps, treatment fidelity was ensured)
4. Results (graph or figure, summary statistics that establish improved student outcomes)

Adequacy of Visual Inspection: The figure presented displays all of the following:
1. At least 3 baseline data points.
2. The baseline is stable or had a trend in an undesirable direction.
3. A sufficient number of treatment data points to evaluate change or growth.

Improved Student Outcomes: The case summary provides a calculation of goal attainment scaling and at least one other summary statistic. Summary statistics support favorable "value added" (see Appendix B for the design and calculation of summary statistics):
1. There is a visible change in trend, level, or variability in the desirable direction across independent phases.
2. Goal Attainment Scaling (GAS)
3. Percent Non-Overlapping Data (PND) and/or Effect Size (ES)

A copy of the case study rubric is displayed in Appendix C

In order to accomplish the evaluation component, all electronically submitted VAPs will be forwarded to a regional evaluation team that consists of university faculty or designees. The State Department of Education will contract the services of each of the state's school psychology programs for training and evaluation purposes. These regional evaluation teams will be responsible for local inservice training, evaluation of VAPs, awarding certificates for approved VAPs and continuing education units (CEU), and communicating all actions to the Department of Education.

Recommendation 3. Renewal of state certification of school psychologists is linked to Value Added Portfolios (VAP).

There are currently various methods to obtain initial DOE certification of school psychologists. The two most commonly used are:

1. Graduation from NASP-approved SSP or PhD program
   - or-
2. National Certification by NASP

Initial state certification will be maintained in this proposal. Renewal of certification currently requires 90 CEUs over five years or national certification by NASP (which has its own more stringent CEU requirements of 75 CEUs over three years). In order to formally adopt the VAP as a central element of professional development, it is recommended that continuing education credits in the amount of 3 CEUs per approved VAP be accepted among the 90 CEUs required by the state for renewal of certification. Thus, three VAPs per year will be submitted by school psychologists to the Department of Education each year, and the applicant will be awarded 3 CEUs along with a Certificate of Completion of Self Study by their regional university. The LSPA/LDOE task force will immediately initiate an agreement with NASP to allow these CEUs as an acceptable credit allowance for national certification.

With these changes, renewal of state certification every five years will require the following:
1. 90 CEUs over five years, including 45 CEUs for 15 Value Added Portfolios
   -or -
1. National certification by NASP **and** 15 approved Value Added Portfolios

   Specialization: VAPs also provide evidence of specialization in each of the 13 disability categories provided in Bulletin 1508. To meet criteria as an area of specialization, the applicant's VAP pool must include five (i.e., an average of one per year) successful case outcomes in one or more of the following categories. Specialty areas do not require additional case studies; the designation simply specifies those disability categories in which the applicant has demonstrated a consistent record of professional practice.

§701	Autism
§703	Deaf - Blindness
§705	Developmental Delay
§707	Emotional Disturbance
§709	Hearing Impairment
§711	Mental Disability
§713	Multiple Disabilities
§715	Orthopedic Impairment
§717	Other Health Impairment
§719	Specific Learning Disability
§721	Speech or Language Impairment
§723	Traumatic Brain Injury
§725	Visual Impairment

Specialty areas will be designated on each applicant's state certification certificate.

<u>Recommendation 4</u>. The value added portfolio (VAP) model is field-tested in targeted school districts prior to state-wide adoption.

State-wide adoption of this proposed model will be administered after a series of graduated field tests are conducted. During <u>Year One</u>, data collection and evaluation will be conducted for practicum sites among the specialist level training programs in the state. During this initial year, selected school districts will be trained in the VAP model and will collect preliminary cases. During <u>Year Two</u>, state-wide adoption of the VAP model for all school psychology interns will be adopted, as well as two additional school districts. During <u>Year Three</u>, state-wide adoption of the VAP model for all practicing school psychologists will be fully implemented.

Bibliography

Barnett, D. W., Daly III, E. J., Hampshire, E. M., Hines, N. R., Maples, K. A., Ostrom, J. K., Van Buren, A. E. (1999). Meeting performance-based training demands: Accountability in an intervention-based practicum. *School Psychology Quarterly, 14,* 357-379.

Bonner, M., & Barnett, D. W. (2004). Intervention-based school psychological services: Training for child-level accountability; preparing for program-level accountability. *Journal of School Psychology, 42,* 23-43.

Kratochwill, T. R., Elliott, S. R., & Busse, R. T. (1995). Behavior consultation: A five year evaluation of consultant and client outcomes. *School Psychology Quarterly, 10,* 87-117.

Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27,* 138-145.

Morrison, J. Q., Graden, J. L., Barnett, D. W. (2007). Steps to evaluating a statewide internship program: Model, trainee, and student outcomes. *Psychology in the Schools, 46,* 989-1005.

APPENDIX A
DEFINITIONS AND EXAMPLES OF SERVICE DELIVERY DOMAINS

CAT1: A universal program that targets a cognitive/academic outcome. This domain includes any strategy that targets a cognitive process directly related to academic skills, such as phonological processing, automaticity, selective attention, interpretation, or planning. These processes may be operationalized in terms of academic production, rate, or accuracy. A second feature is that intervention supports have the potential to strengthen general instruction for *all* children in a school. A distinguishing characteristic of interventions in this domain is *capacity building*, meaning that changes in systems (e.g., school-wide), general instruction (e.g., teaching routines), or management (i.e., data-based decision making) increase the probability that future problems will be prevented. Strategies for evaluation may include:

- *Systems Change:* The effect of an SBLC training on pneumonic strategies is evaluated in terms of changes in test performance for the first relevant case encountered by the team.

- *General Instruction:* The effect of behavioral consultation with a second grade teacher on goal setting and public posting is evaluated in terms of changes in class-wide weekly spelling grades.

- *Management:* The effect of a new universal screening procedure for comprehension is evaluated in terms of changes in the average maze score across all children in the school.

For any of these strategies, the method of evaluation may be varying samples of the target population (individual, group, or school-wide).

SAT1: A universal program that targets a social/affective outcome. This domain includes any strategy that targets a social or affective behavior or outcome, such as percentage of children reporting bullying, rate of office referrals, or frequency of prosocial behaviors. A second feature is that intervention supports have the potential to strengthen behavioral outcomes for *all* children in a school. A distinguishing characteristic of interventions in this domain is *capacity building*, meaning that changes in systems (e.g., school-wide), general instruction (e.g., teaching routines), or management (i.e., data-based decision making) increase the probability that future problems are prevented. Strategies for evaluation may include:

- *Systems Change:* The effect of an SBLC training on self-mutilation is evaluated in terms of changes in self-injury for the first relevant case encountered by the team.

- *General Instruction:* The effect of class-wide social scripts with an eighth grade teacher is evaluated in terms of changes in the teacher's rate of office referrals.

- *Management:* The effect of a new check in/check out procedure for truancy violators is evaluated in terms of changes in absenteeism across all children in the school.

For any of these strategies, the method of evaluation may be varying samples of the target population (individual, group, or school-wide).

CAT2: A selective-group remediation program that targets a cognitive/academic outcome. This domain includes any strategy that targets a cognitive process directly related to academic skills, such as phonological processing, automaticity, selective attention, interpretation, or planning. A second feature is that intervention supports have the potential to remediate a common target using a standard protocol

for all children in a group. A distinguishing characteristic of interventions in this domain is *selective-group*, meaning that a small group of children are matched to the treatment. Strategies for evaluation may include:

- The effect of phonemic awareness training is evaluated in terms of changes in the average DIBELS scores of a targeted group of first grade children.

- The effect of an afterschool tutoring program is evaluated in terms of changes in the average math scores of three participants with 100% attendance record.

- The effect of behavioral consultation with a second grade teacher on self-monitoring is evaluated in terms of changes in homework accuracy for the child with the lowest homework grade.

For any of these strategies, the method of evaluation may be varying samples of the targeted group (one individual or a group average).

SAT2: A selective-group remediation program that targets a social/affective outcome. This domain includes any strategy that targets a social or affective outcome, such as percentage of children reporting bullying, rate of office referrals, or frequency of prosocial behaviors. A second feature is that intervention supports have the potential to remediate a common target using a standard protocol for all children in a group. A distinguishing characteristic of interventions in this domain is *selective-group*, meaning that a small group of children are matched to the treatment. Strategies for evaluation may include:

- The effect of a peer mentorship program is evaluated in terms of changes in the average direct behavior ratings of a targeted group of ninth grade children.

- The effect of a coping skills group is evaluated in terms of changes in the average rate of teacher-reported aggression for three participants who completed the program.

- The effect of behavioral consultation with a sixth grade teacher on homework completion is evaluated in terms of changes in the percentage of homework completion for the lowest performing child in the class. .

For any of these strategies, the method of evaluation may be varying samples of the targeted group (one individual or a group average).

CAT3: An individualized intervention that targets a cognitive/academic outcome. This domain includes any strategy that targets a cognitive process directly related to academic skills, such as phonological processing, automaticity, selective attention, interpretation, or planning. A second feature is that intervention supports are delivered to a single child. A distinguishing characteristic of interventions in this domain is *specialized instruction*, meaning that the treatment is intensive, delivered by a highly qualified specialist, and is tailored to the specific instructional needs of an individual child. Strategies for evaluation may include:

- The effect of an intensive and individualized strategy training is evaluated in terms of changes in the comprehension scores of a child referred for special education services.

- The effect of an intensive and individualized phonemic awareness program is evaluated in terms of changes in initial sound fluency for a preschool child.

- The effect of an intensive and individualized repeated reading program is evaluated in terms of oral reading fluency growth for a second grade child on an I.E.P.

For any of these strategies, the method of evaluation is analysis of individual growth or performance.

SAT3: An individualized intervention that targets a social/affective outcome. This domain includes any strategy that targets a social or affective outcome, such as percentage of children reporting bullying, rate of office referrals, or frequency of prosocial behaviors. A second feature is that intervention supports are delivered to a single child. A distinguishing characteristic of interventions in this domain is *specialized instruction*, meaning that the treatment is intensive, delivered by a highly qualified specialist, and is tailored to the specific behavioral needs of an individual child. Strategies for evaluation may include:

- The effect of an intensive and individualized behavior management program is evaluated in terms of changes in peer-directed aggression for a child referred for special education services.

- The effect of an intensive and individualized peer mediation program is evaluated in terms of changes in office referrals for a preschool child.

- The effect of an intensive and individualized counseling program is evaluated in terms of changes in direct behavior ratings for a sixth grade child on an I.E.P.

For any of these strategies, the method of evaluation is analysis of individual growth or performance.

APPENDIX B
GUIDELINES FOR GATHERING IMPACT DATA
FOR INTERVENTION CASES

(Adapted from *Guidelines for Gathering Impact Data*
*for the Ohio School Psychology Internship Evaluation, 2007)*

All intervention case reports must provide goal attainment scaling and at least one other sources of impact data. Step-by-step guidelines for PND, GAS, and ES are provided below.

Step-by-Step Guide for Calculating Percentage of Non-overlapping Data

For an intervention designed to increase the target behavior, determine the percentage of the intervention data points that fall above the highest baseline data point. In the example that follows, all intervention points except one are above the highest baseline data point. Therefore 9 of the 10 intervention data points do not overlap the baseline, and PND = 90%.
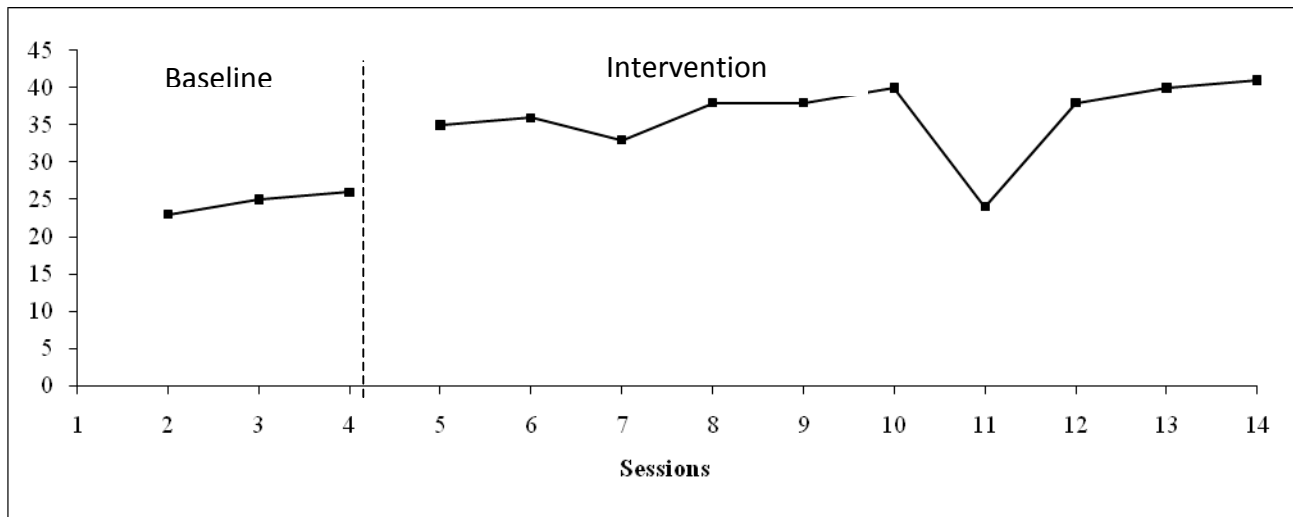
For an intervention designed to decrease the target behavior, determine the percentage of the intervention data points that fall below the lowest baseline data point. In the example that follows, the first four intervention data points are below the lowest baseline data point, as well as the seventh and eighth intervention data points. Therefore, 6 of the 10 intervention data points do not overlap baseline, and PND = 60%.
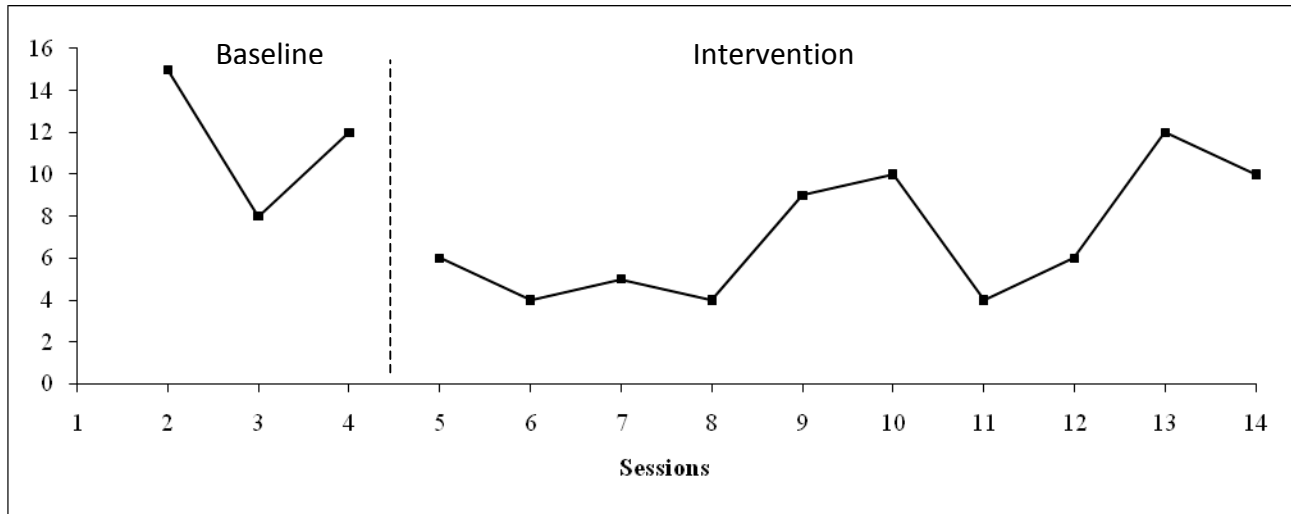
Considerations When Using PND

❑ PND should not be used if there are extreme scores, such as a 0 during baseline for a target behavior you want to decrease, or an extremely high value during baseline for a target behavior you want to increase.
❑ PND scores at or above 70% indicate the intervention is effective; scores between 50% - 69% show moderate intervention effects. PND scores below 50% would be considered ineffective.

Step-by-Step Guide to Developing and Scaling Goals
Using Goal Attainment Scaling

Goal attainment scaling (GAS) provides a subjective evaluation of the success or failure of an intervention, and can be used to (a) provide an overall subjective rating of performance changes that are represented by the primary outcome measure, such as oral reading fluency, tallies of problem behavior, or (b) provide an overall subjective rating of some other potential intervention outcome that was not directly measured. Due to its subjective nature, the GAS provides both an opportunity to assess dimensions and domains of performance that are not easily quantified, while also introducing the risk of rater bias. Therefore, definitions of rating anchors and the evaluation of child performance should be the product of a team, rather than one individual. There are four steps to developing a goal attainment scale:

STEP ONE

Define a "0" anchor by describing current levels of functioning. If the primary outcome measure is the most acceptable indicator of progress, then the anchor statement should reflect the range of scores or a growth rate that was observed during baseline. If the primary outcome measure does not assess an important domain or dimension of performance that is equally important but more difficult to measure

directly, then the team should create a statement of performance that complements the primary outcome measure.

STEP TWO

Define the "+1" anchor by describing a reasonable goal of the intervention, keeping the following in mind:

- A reasonable goal should be based upon what the student will likely achieve by the end of the intervention. Local or literature-based standards are available for most academic and behavioral outcomes,
- Reasonable goals should take into consideration the child's current levels, the usual outcomes of this intervention, the amount of resources and time planned for intervention, and the skills of the intervention specialist/change agent. For example, it may be unrealistic to expect a child to reach the same level as his or her peers, but entirely appropriate to target a rate of progress (e.g., weekly improvement) that is equivalent to peers or literature-based standards.
- A reasonable goal should be acceptable to the teacher, parent, and student.
- Reasonable goals should focus, in most instances, on targeting positive replacement behaviors *rather than decreases in problem behavior*.

STEP THREE

Define the "-1" anchor by describing the lowest or worst level of performance observed during baseline. This score should represent an unacceptable level of performance, but would also be expected if there were no intervention supports in place.

STEP FOUR

Define the "+2" and "-2" anchors. A "+2" score should reflect an ambitious goal and *greater than expected* outcome. A "-2" score should reflect a previously unobserved level of performance that would place the child at the highest risk for poor social or academic consequences.

The design of GAS anchors and the evaluation of child performance should be based on collaborative decision making among all members of a problem solving team, such as the teacher, child, school psychologist, and/or parent. Two examples of GAS scales are provided below.

GOAL ATTAINMENT SCALE (Example 1)

For an intervention designed to increase reading comprehension, the child was administered a comprehension intervention while progress was monitored weekly using maze passages. A figure indicates that during baseline the child was completing a grade-level maze passage with 50% - 60% accuracy. The goal set by the SBLC team was to improve maze performance to 80% on three consecutive weeks, but to also improve the child's classroom performance on weekly language arts tests and classroom/homework assignments. Goal attainment scaling was designed to complement the evaluation of maze performance. Thus, the following behavioral anchors were developed for evaluating goal attainment.

Much worse -2	*Marcello's grades on both language arts tests <u>and</u> assignments have worsened since the intervention began.*
Slightly worse -1	*Marcello's grades on either language arts tests or related assignments have worsened since the intervention began.*
No change 0	*Marcello's grades on language arts tests and assignments have not changed since the intervention began.*

Slightly improved +1	*Marcello's grades on either language arts tests or assignments have improved since the intervention began.*
Much Improved +2	*Marcello's grades on both language arts tests <u>and</u> assignments have improved since the intervention began.*

Each extreme level (-2, +2) represents the outcome that might be expected to occur in 5% to 10% of similar at-risk students.

GOAL ATTAINMENT SCALE (Example 2)

For an intervention designed to decrease aggression displayed by a fifth grade child, the I.E.P. team used teacher weekly recordings of aggression toward peers and adults to evaluate a Positive Peer Reporting strategy. Goal attainment scaling was used to provide a broader, more global indicator of the child's pro-social skills. The following behavioral anchors were developed for evaluating goal attainment.

Much worse -2	*Charron displays daily episodes of verbal and physical aggression toward peers and adults*
Slightly worse -1	*On some days, Charron displays verbal and physical aggression toward peers and adults*
No change 0	*Charron's aggression and respect for others has not changed since the intervention was begun.*

Slightly improved +1	*On most days, Charron interacts well with others and demonstrates self-control in conflict situations*
Much Improved +2	*Charron develops friendships and exhibits age-appropriate respect for others. Her behavior is not noticeably different from other children.*

Each extreme level (-2, +2) represents the outcome that might be expected to occur in 5% to 10% of similar at-risk students.

Step-by-Step Guide for Calculating Effect Sizes

STEP 1

Calculate the mean of the baseline data points. In the example below, the mean of the three baseline data points (18, 6, 12) is 12.0.

STEP 2

Calculate the mean of the intervention data points. In the example below, the mean of the ten intervention data points (4, 3, 3, 3, 5, 5, 6, 6, 3, 2) is 4.0.

STEP 3

Calculate the standard deviation of the baseline data points. In the example below, the standard deviation of the three baseline data points (18, 6, 12) is 6.0.
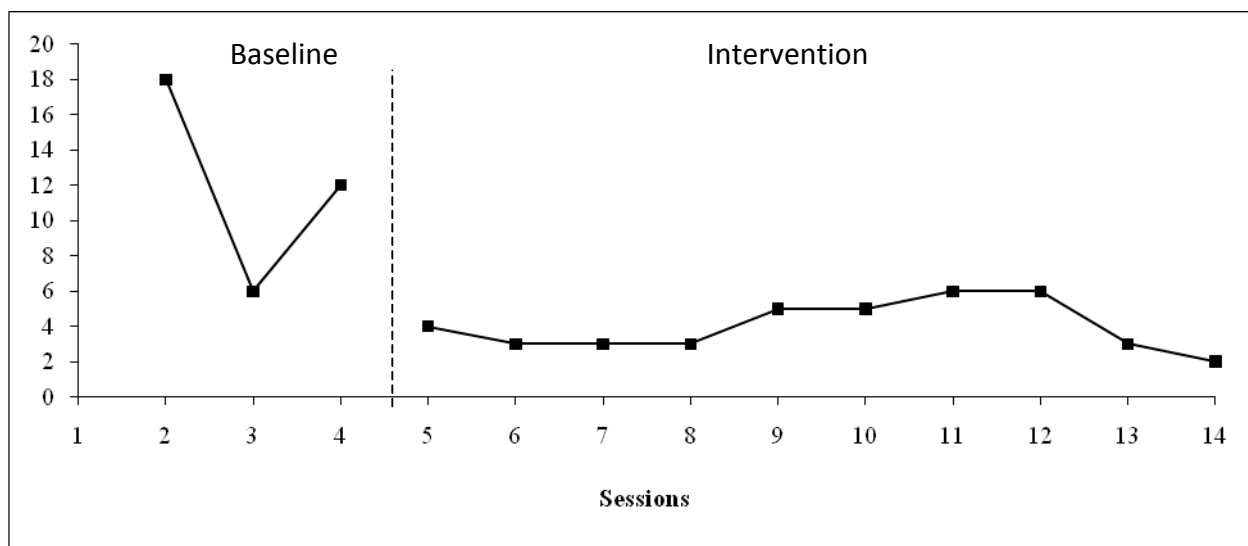
STEP 4

Calculate the difference between the mean of the intervention data and the mean of the baseline data and divide that by the standard deviation of the baseline data.

ES = mean of the intervention data − mean of the baseline data

standard deviation of baseline data

In the example below,

ES = $\frac{4.0 - 12.0}{6.0}$ = -1.33

Considerations When Using ES

- ❑ ES should not be used if the baseline data are limited to one or two points only.
- ❑ ES should not be used if there is no variability in the baseline data (*SD* = 0)
- ❑ Interventions that target increases in behavior should result in a positive ES (e.g., +2.5) while intervention that target decreases in behavior should result in a negative ES (e.g., - 2.5)
- ❑ ES is equivalent to standard deviation units, thus any effect size greater than -3.0 or +3.0 (representing 3 SD) should be reported as an ES of -3.0 or +3.0. In other words, 3 SD is the "ceiling."
- ❑ An ES of +.67 or greater is considered effective; an ES between +.33 -.67 is marginal or moderate; an ES below +.33 is considered small.

APPENDIX C
CASE STUDY SCORING RUBRIC
(Adapted from NASP NCSP Case Study Rubric)

	Needs Development	Standard	Passing
1.1	☐ Child characteristics (age, grade, demographics) are not included in summary.	Child Characteristics	☐ Child characteristics are clearly described.
1.2	☐ The problem is not adequately defined: a. ☐ Not measurable b. ☐ No peer/normative comparison data c. ☐ Missing goal data or goals not appropriate/evidence-based	Problem Identification Identifies an observable and measureable target for intervention. Collaboratively developed problem statements capture multiple perspectives and improve intervention acceptability.	☐ The problem is defined in terms of both current and desired levels of performance, including comparisons to peer performance and/or local/national benchmarks. [1] ☐ An intervention goal stating both summative and formative measures and how these measures were selected is included. [2]
1.3	☐ Intervention selection is not based upon hypothesis testing. a. ☐ Intervention is based upon diagnostic category, not empirical evidence of effectiveness with this student. b. ☐ Hypotheses stated were not testable/tested.	Problem Analysis and Intervention Design Includes functional assessment of behavior/academic performance. Hypothesis generation and testing. Hypothesis testing is **the** critical mechanism for linking assessment to intervention when creating an individualized instructional intervention.	☐ One or more hypotheses that explain the conditions in which the problem is likely to occur or result from are stated. At least one hypothesis must address performance/skills deficits. ☐ Hypotheses are restated as predictions for outcomes based upon selected intervention and tested forming a basis for intervention selection.
1.4	☐ There is no graph and/or summary of results or graphed data was not used to inform intervention implementation (formative evaluation)	Results	☐ There is a graph displaying baseline, goal, and treatment data that is summarized.
2.1	☐ There are too few baseline data points.	Adequacy of Baseline Data	☐ There are at least three baseline data points.
2.2	☐ The baseline data are not stable or exhibit a trend in the desirable direction.	Adequacy of Baseline Trend	☐ The baseline data are stable or had a trend in the undesirable direction.
2.3	☐ There are too few treatment data points	Adequacy of Treatment Trend	☐ There are at least three treatment data points.

3.1	☐ There is no visible change in trend, level, or variability.	Visual Inspection	☐ There is a visible change in trend, level, or variability in the desired direction.
3.2	☐ GAS suggests an ineffective intervention.	Goal Attainment Scaling	☐ GAS indicates an effective intervention.
3.3	☐ PND suggests an ineffective intervention ☐ ES suggests an ineffective intervention	Percentage Non-Overlapping Data Effect Size	☐ PND indicates an effective intervention. ☐ ES indicates an effective intervention.

.