Internships play a pivotal role in the future of school psychology, and internship training is influenced by disciplinary trends including (a) an increased focus on accountability and demonstrating the positive impact of trainees’ practices on student outcomes and (b) growing emphasis on scientifically based and empirically based practice. These influences also include greater sophistication in decision making as well as different ways of outcome-focused decision making. System-wide approaches such as Positive Behavior Support (PBS) and response to intervention (RTI) require high training demands to develop intervention-related decision and accountability skills. Programmatically, accountability systems to meet the training needs of intervention services require scaling up. This article describes a multiyear developmental project bringing diverse universities together to institutionalize a statewide system of accountability. We share data from 3 years (N of interns = 278) and suggest recommendations for outcome-oriented evaluation for other universities, agencies, school psychological services delivery units, or states.

Forces that drive accountability movements in education and psychology have a great impact on the field of school psychology. Although accountability has played a prominent role in the educational reform efforts of the past 50 years (Linn, 2000), the No Child Left Behind Act (NCLB) (PL 107-110, 2001) intensified accountability efforts by mandating sanctions for schools and districts that do not meet academic performance standards. Two key positions of NCLB have influenced the practice of school psychology in particular: stronger accountability for results and proven educational methods. School psychologists will need to examine professional functions and quality of services, not only by ratios of professionals to students or other broad indicators such as number of students tested or “served” (cf., Goodman & Young, 2006). Future efforts to examine the impact of the support services will be guided by what school psychologists do and to what effect.

In school psychology, there likewise has been an increasing focus on accountability and on promoting services that have a demonstrated positive impact on students. Training standards of the National Association of School Psychologists (NASP, 2000b) reflect the growing focus on accountability in accreditation and training standards and include emphasis on performance-based program accountability. The NASP Guidelines for professional practice and services (NASP, 2000a) encourage school psychologists and school psychological service units to evaluate services provided and their impact on students and families. In the 2002 Conference on the Future of School Psychology, all key groups representing the profession promoted several broad themes to guide the future (Dawson et al., 2004). These themes included a focus on systems-level practices, prevention and early intervention, evidence-based interventions, and assessment linked to interventions and accountability. The most recent NASP Blueprint for Training and Practice (NASP, 2006) emphasized a focus on promoting positive outcomes for children, data-based decision making, and systems-based service delivery.

Accountability is emphasized in the evidence-based practice movement (Cowan & Skalski, 2008). The Task Force on Evidence-Based Interventions in School Psychology served the dual
purpose of (a) examining and disseminating the knowledge base on prevention and intervention programs or approaches for children, youth, and families that demonstrated empirical support for application in the school and community, and (b) extending the knowledge base through facilitating sound research methodologies, technologies, and innovations (Kratochwill & Stoiber, 2002).

**Accountability in School Psychology**

School psychologists have a critical role in promoting positive student outcomes and evidence-based practices, but only if they have the knowledge and skills needed to meet the current demands of the profession. System-wide approaches that focus on consultation, problem-solving, contextual data sources, and interventions at school, class, and individual levels pose training challenges to the school psychology profession to reach professional levels of preparation (Barnett, Daly, Jones, & Lentz, 2004; Batsche et al., 2005; Gresham et al., 2005; O'Shaughnessy, Lane, Gresham, & Beebe-Frankenberger, 2003). Thus training missions of universities and internship programs are extensive and evolving to meet the needs of an ever-advancing profession.

**Accountability in School Psychology Training Programs**

Accountability practices for measuring student outcomes must be an integral part of school psychology training programs to promote their use in educational settings. Two facets of accountability for school psychology trainees include (a) using accountability methods for intervention services to help evaluate student outcomes, and (b) using the data from outcomes for program-level accountability in meeting training objectives (Barnett et al., 1999). Within the literature on training programs, only four published reports were found that include data on child outcomes directly related to school psychology training procedures (Barnett et al., 1999; Bonner & Barnett, 2004; Kratochwill, Elliott, & Busse, 1995; Sheridan, 1992). The training targets have been field practicum and consultation skills. The impact of internship experiences on child outcomes has not been systematically evaluated. Basic evaluation methods, expanded later, use facets of single case research designs (e.g., baseline, intervention; referred to as an A-B design). The data from A-B designs can be used for instructional and intervention decision making based on level and trend and as accountability data that can be summed over discrete interventions in a coherent series or group of interventions for evaluation purposes. Internal validity is not demonstrated with an A-B design because other possible factors related to change cannot be ruled out. However, to the degree that empirically established interventions are used in school services, accountability designs may be used for service delivery evaluation and the A-B design is its “workhorse” (Bloom, Fischer, & Orme, 2005, p. 353). In addition, other standard methods of outcome determination can be used based on data from replicated A-B designs, such as effect size (ES), percentage of non-overlapping data (PND), and Goal Attainment Scaling (GAS).

**Internship Training in School Psychology.** Standards for internship training are addressed in training standards of the NASP (2000b). The standards outline the domains and structures of school psychology training, including the internship, with standards addressing the alignment of the internship with training program objectives, number of hours and diversity of field-based experiences, requirements for field supervisor credentials and supervision commitment, and the requirement that training programs develop a plan for performance-based assessment and program accountability. The performance-based assessment and accountability standard (Standard 4) states that “a key aspect of program accountability is the assessment of the knowledge and capabilities of school psychology candidates and of the positive impact that interns and graduates have on services to children, youth, families, and other consumers” (p. 19). The specific performance-based standards include using systematic, valid procedures to evaluate and improve the quality of the training program (Standard 4.1); applying performance-based assessments of trainees academic and professional
competencies and professional work characteristics (Standard 4.2); and demonstrating that trainees can integrate and apply knowledge and skills to provide a “comprehensive range of services evidenced by measurable positive impact” on students.

The Ohio Internship Program

Ohio has a significant history of supporting the training of school psychologists through a partnership with the Ohio Department of Education (ODE), Office for Exceptional Children (OEC). For more than 40 years, ODE has funded Ohio school psychology internships. Since the late 1970s, all Ohio school psychology training programs have collaborated with the ODE through an Inter-University Council (IUC) for School Psychology formed for the purpose of promoting school psychology training aligned with national standards and also focused on addressing statewide internship issues. The IUC is composed of faculty from the state’s school psychology programs (current number of programs is nine) as well as representation from the ODE and the Ohio School Psychologists Association (OSPA). The Council meets three times per year to guide training efforts, with focused subcommittees meeting as needed. The IUC has adopted formal policies and guidelines for the Ohio internship and has committed to alignment with NASP training standards. Key past activities have included adopting The Ohio Internship Program Manual (ODE, 2006) that uses the NASP training standards as a foundation and adds specific Ohio requirements to establish collaborative training and practice priorities across the state. In the 1980s, prior to all Ohio programs being reviewed by NASP, the IUC engaged in an in-depth process of peer review (including on-site visits) that supported program development toward NASP subsequent program approval at all institutions. Since 2004, the Ohio Internship Program has operated as a special program through the ODE, and IUC is accountable for demonstrating important training outcomes that are aligned with ODE priorities consistent with recent school mandates and legislation as well as school psychology reforms.

Ohio provides funding for internships through the ODE to school districts/educational service centers that are approved by the state on criteria adopted by IUC. Chief among these criteria are alignment with NASP training standards (NASP, 2000b), including commitment to the internship as a training program that supports the training mission of the Ohio Internship Program as well as the specific training goals and philosophy of individual training programs. Training sites are submitted for approval by IUC programs, evidencing the partnership between local school districts and the state. There is collaboration and sharing among these approved sites across training programs (e.g., within regions, a site may have interns from more than one university program; programs may agree to provide an intern from another region to complete an internship in an approved site). Training sites are reviewed on a 3-year cycle for reapproval. Approval includes review by the IUC as well as final approval by the state. The IUC also has adopted policies governing the number and ratio of interns to training programs that are aligned with NASP standards. All Ohio interns accepting an internship in the Ohio Program sign a commitment to remain in the state for a minimum of 1 year following completion of the internship, to assure that the Ohio training commitment is contributing to the significant need in the state for comprehensively trained school psychologists. The internship program and site approval procedures are detailed in The Ohio Internship Program Manual (ODE, 2006).

Internship field supervisors are selected within a school district that has attained site approval status based on the following minimal requirement: “at least two years of successful full-time experience as a school psychologist, one of which must be as an employee in the present school system” (ODE, 2006, p. 14). University faculty work collaboratively with district administrators to determine which school psychologists would be willing and best able to serve as effective internship
field supervisors. Internship supervisors are not compensated for serving in this capacity. Beyond assuring that the minimal requirements for supervisors are met, university faculty were not required to obtain additional information regarding the supervisors’ professional training and experiences.

In summary, the Ohio Internship Program has a significant history as well as future that, at least in part, may be due to evaluation efforts as well as considerable teamwork that address state and national initiatives consistent with significant priorities in school psychology. Recent emphases in both accountability for services and shifts toward evidence-based intervention decisions has led to a proactive evaluation of state internships with regard to service outcomes for schools and students.

**METHOD**

**Purpose of the Evaluation**

The purpose of this statewide evaluation was to assess the effectiveness and impact of the Ohio Internship Program. There were two components to the outcome-oriented evaluation: (a) demonstrating effectiveness in terms of interns’ skill attainment and competence in key domains, and (b) demonstrating the impact of interns’ activities and practices on student outcomes. Because of the commitment from the state to training, there was a desire to document the impact of interns’ activities in prioritized areas. This effectiveness and impact evaluation was consistent with the outcome orientation of NASP training standards, which require programs to document the impact of trainees’ activities on the students they serve. The evaluation plan was developed by an IUC subcommittee, with ODE input and approval, and approval by the IUC membership, which includes trainers from all Ohio programs.

**Evaluation Questions**

Evaluation questions were developed and agreed upon by the members of the IUC as relevant and meaningful across all nine school psychology programs. There were two major questions. The first question addressed the degree to which school psychology interns demonstrated an increase in professional skills and competencies during their Ohio Internship. The second question focused on the impact of interns’ practices on student outcomes, in terms of (a) number of students served across tiers (universal—Tier 1; supplemental intervention—Tier 2; and intensive, individualized intervention—Tier 3) as well as demographic information on students served, and (b) student outcomes in terms of goal attainment and summary statistics reflecting effectiveness of intervention strategies.

**Target Population: Ohio School Psychology Interns**

All Ohio interns entering the Internship Program have been determined by their training programs to be eligible for the Internship (e.g., completed all coursework and applied experience, demonstrated readiness for internship). Documentation for all recommended interns is submitted for a final review by a subcommittee of the IUC, which assures that all training criteria for temporary licensure, including required criminal background checks have been completed. The recommended interns are submitted to ODE for final approval for temporary licensure (1 year) as intern school psychologists.

The participants in this evaluation included 278 school psychology graduate student interns who participated in the Ohio Internship Program in 2004–2005 ($n = 97$), 2005–2006 ($n = 95$), and 2006–2007 ($n = 86$). All interns completed their graduate coursework at one of the nine university training programs in Ohio (all programs were NASP approved). All but one of the interns (99.6%) provided written consent to have their internship data included in the statewide evaluation report. (This study was submitted to the Committee for the Protection of Human Subjects at the principal investigator’s university. Additional consents and assents were not judged to be warranted for the

*Psychology in the Schools*  DOI: 10.1002/pits
purposes of this statewide evaluation). Interpretable data were received by slightly fewer interns for various reasons (e.g., missing permissions, data), but the number of interns providing data was fairly consistent and improved across all three years: 2004–2005 ($N = 86; 88.7\%$), 2005–2006 ($N = 92; 96.8\%$), and 2006–2007 ($N = 86; 100\%$).

Ohio interns were supervised by one or more university-approved field supervisors. University-level supervision also was provided consistent with NASP standards.

**Evaluation Design**

This evaluative study employed descriptive research methods to summarize quantitative outcomes of the Ohio Internship Program. To assess development of interns’ skills and competencies during the internship, university-developed rating scales were completed by internship field supervisors at the beginning, at the midpoint, and at the end of the internship. To assess impact, interns were asked to provide data regarding the students they served based on the professional practice logs they were required to maintain throughout the school year, as prescribed by The Ohio Internship Program Manual (ODE, 2006). A subset of interns across three training programs also provided student outcome data in relation to specific, measurable, and individualized student goals and graphed data. These student outcome data were submitted directly to their university supervisors as part of the interns’ portfolio.

**Procedures**

**Effectiveness of the Internship on Intern Skill and Competency Development.** Overall effectiveness of the Ohio Internship Program was evaluated summatively through ratings of attainment of intern competencies. Each university training program had developed an intern competency evaluation instrument based on (a) the program’s training goals, (b) NASP training standards and other accreditation guidelines, as well as key training and practice documents (e.g., NASP Blueprint); and (c) pertinent federal, state, and local laws, regulations, and procedures. The internship competencies evaluation instruments emphasized strong documentation and accountability for professional skills. Because of the outcome orientation of the NASP training standards and Ohio programs’ alignment with these standards, all state school psychology training programs had items that evaluated interns’ skills on important domains related to intervention-oriented practice. These individual competency evaluation instruments were reviewed by the evaluation subcommittee to identify a common core of items that evaluated interns’ competencies in providing services to students to improve academic and behavioral outcomes. Individual programs reviewed item selection to affirm that the correct items were identified. This process resulted in a common set of items to summarize intern effectiveness on student outcomes (See Appendix). The lack of a single instrument for use across all of the university training programs is a limitation of this model.

**Outcome Evaluation: Breadth of Service Delivery and Impact on Students Served.** Fundamental to this aspect of the outcome evaluation is alignment with Ohio’s commitment to a schoolwide focus on improving outcomes for all students, and closing achievement gaps, through a comprehensive system of effective instruction and intervention. Thus, the evaluation focused on these aspects of outcomes: (a) number of students served by interns across tiers, including Tier 1 (universal/system level practices such as Positive Behavior Support (PBS) planning and universal screening for instructional decision making), Tier 2 (supplemental/targeted intervention), and Tier 3 (intensive/individualized interventions); (b) demographic description of the students served at Tier 2 and Tier 3, reported by NCLB categories (i.e., race/ethnicity, disability status, economically disadvantaged status, and Limited English Proficiency [LEP] status) as well as gender; and (c) measurement of student outcomes. For this final outcome, interns were required to select at least two academic interventions (Tier 2 and Tier 3) and three behavior interventions (Tier 1, 2, and 3) in which they were
meaningfully involved (a type of *purposive* sampling; Patton, 1990). Tier 1 academic practices were recommended by the IUC for inclusion, but not required prior to 2007–2008, due to concerns that an intern might not have sufficient opportunity to be involved at a meaningful level with system-level practices until there was a higher level of implementation in their districts. In subsequent years, Tier 1 academic interventions were included in the statewide evaluation. Interventions were described at program levels and were not described further for the state-level evaluation.

For each of the five interventions identified as exemplars of their professional practice, interns were asked to summarize the outcome demonstrated by the student or students using GAS. In 2006–2007, three training programs volunteered for their interns to contribute additional outcome and impact data, which included ES estimates and PND.

**Outcome Measures**

**Intern Competencies Evaluation Instrument.** Field-based internship supervisors were required to evaluate the performance of the interns they supervised and provide ratings on three occasions throughout the school year: Fall, Winter, and Spring on each university’s individual intern evaluation instrument. A common scale was used across these instruments, structured as a 4-point categorical rating scale where 4 = Mastery, 3 = Satisfactory, 2 = Developing, and 1 = Unsatisfactory. Universities identified critical domains for the purposes of this evaluation: (a) use of assessment in a problem-solving context, (b) knowledge and skill in consultative problem-solving skills, (c) use of empirically based academic intervention strategies, (d) use of empirically based behavioral intervention strategies, (e) use of data to monitor progress, and (f) competency in conducting professional development activities.

**Student Demographic Data.** Interns were asked to record as separate counts the number of students for whom they provided Tier 1–3 services. Tier 1 services were defined as those supporting effective system practices (schoolwide and classwide implementation), such as participating in assisting with activities for universal screening and decision making based on data, serving on decision-making teams for planning effective core practices such as PBS. Tier 2 services were defined as supplemental/targeted intervention, such as assisting in planning, delivering, and evaluating research-based academic and behavioral interventions. Tier 3 services were defined as intensive/individualized interventions, such as assessment for intervention; planning, implementation, and outcome evaluation of research-based interventions; and crisis counseling.

Interns also provided demographic data for students served at Tier 2 or Tier 3 at each of their schools. The demographic data requested included students’ gender, race/ethnicity, disability status, economically disadvantaged status, and LEP status.

**Methods for Outcome Determination**

Achieving program-level accountability requires an objective method for assessing outcomes across multiple interventions (e.g., Scruggs & Mastropieri, 1998). The synthesis of intervention outcomes as well as the appropriateness of various approaches for summarizing single subject research has received considerable attention in recent years (Olive & Smith, 2005; Parker & Brossart, 2003; Scruggs & Mastropieri, 2001). GAS was the primary method used for summarizing intervention outcomes for students served by school psychology interns. As a supplement to the GAS process, two additional summary statistics were calculated among interns from three university training programs to measure the effects of an intervention provided by the interns: the PND and ES.

**GAS.** The GAS process involved the development of a 5-point scale for measuring goal attainment as outlined by Kiresuk, Smith, and Cardillo (1994). In this study, "Expected Level of
Outcome” was replaced with “No Change” to better represent students’ responses to the intervention (RTI). Thus, positive ratings reflected a positive change in the target, and negative ratings reflected a change in an undesired direction for the target. The other scale anchors remained the same: “Somewhat More than Expected,” “Somewhat Less than Expected,” “Much More than Expected,” and “Much Less than Expected.”

Reviews of the reliability and validity of many applications of GAS procedures are available in Cardillo and Smith (1994) and Smith and Cardillo (1994), respectively. Studies that used a 5-point scale (similar to the approach used herein) reported inter-rater reliability indices between .87 and .93 (as cited in Cardillo & Smith, 1994). Test–retest reliability also was acceptable (e.g., correlation of \( r = .84 \) over a two- to three-week period; see studies reported in Cardillo & Smith, 1994). In school settings, the use of GAS methodology has been demonstrated to be of significant value in the evaluation of intervention-based change, and is “a more accurate estimate than any other measure” (Sladeczek, Elliott, Kratochwill, Robertson-Mjaanes, & Stoiber, 2001, p.52). GAS validity evidence includes analyses of many types of intervention outcomes including school-based interventions (see Kratochwill et al., 1995). GAS has been found to be responsive to measuring diverse functional goals across services and sensitive to measuring intervention-induced change, making it a strong outcome measure for groups of students in which the rate of progress varies (MacKay, McCool, Cheseldine, & McCartney, 1993). A summary of the research regarding the utility and acceptability of GAS for measuring students’ progress can be found in Roach and Elliott (2005).

PND. Calculating the PND involves counting the number of intervention data points that exceed the highest baseline point (for studies seeking to increase a target behavior) or counting the number of intervention data points lower than the lowest baseline point (for studies seeking to decrease a target behavior). The number of non-overlapping data points is then divided by the total number of intervention points to obtain the PND. PND has been found to produce a summary statistic that is consistent with the outcomes obtained through visual analysis of individual participant graphs (Olive & Smith, 2005). PND should not be calculated when a baseline data point of zero is present in decreasing behavior studies or an extremely high baseline data point is present in increasing behavior studies (Scruggs & Mastropieri, 1998; Scruggs, Mastropieri, & Casto, 1987).

The use of PND as a summary statistic that is easy to calculate and interpret has wide support in the research literature (Mathur, Kavale, Quinn, Forness, & Rutherford, 1998). Ratings using PND are judged on the following scale: a PND greater than or equal to 90% is considered “Highly Effective,” a PND of 70% to less than 90% is judged as “Moderately Effective,” a PND of 50% to less than 70% is considered “Mildly Effective,” and a PND of less than 50% is rated as “Ineffective” (Scruggs, Mastropieri, Cook, & Escobar, 1986).

ES. There are many ES estimation methods (Busk & Serlin, 1992; Thompson, 2007). ES in this study was calculated as the change in achievement or behavior relative to the baseline (control) standard deviation (\( \text{SD;} \) Busk & Serlin, 1992). As a general guide for outcomes without much specific prior evidence for comparisons, interventions that yield an ES greater than or equal to 0.80 are considered to have a large effect; an ES between 0.50 and 0.79 represents a moderate effect, whereas an ES between 0.20 and 0.49 reflects a small effect.

RESULTS

To What Degree Did School Psychology Interns Demonstrate an Increase in Professional Skills and Competencies During Their Ohio Internship?

Internship supervisor ratings were obtained for 266 school psychology interns across three school years: 2004–2005 (\( N \) of ratings = 90), 2005–2006 (\( N = 90 \)), and 2006–2007 (\( N = 86 \)).
Table 1

<table>
<thead>
<tr>
<th>Domains of Practice</th>
<th>Fall Mean SD</th>
<th>Winter Mean SD</th>
<th>Spring Mean SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Assessment in Problem-Solving Context</td>
<td>2.49</td>
<td>3.23</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.71</td>
<td>0.38</td>
</tr>
<tr>
<td>Consultative Problem-Solving Skills</td>
<td>2.35</td>
<td>3.22</td>
<td>3.86</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>0.56</td>
<td>0.40</td>
</tr>
<tr>
<td>Academic Intervention Strategies</td>
<td>2.35</td>
<td>3.18</td>
<td>3.81</td>
</tr>
<tr>
<td></td>
<td>0.76</td>
<td>0.54</td>
<td>0.40</td>
</tr>
<tr>
<td>Behavioral Intervention Strategies</td>
<td>2.34</td>
<td>3.10</td>
<td>3.74</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.61</td>
<td>0.45</td>
</tr>
<tr>
<td>Use of Data to Monitor Progress</td>
<td>2.34</td>
<td>3.15</td>
<td>3.80</td>
</tr>
<tr>
<td></td>
<td>0.77</td>
<td>0.57</td>
<td>0.43</td>
</tr>
<tr>
<td>Conducting Professional Development Activities</td>
<td>2.15</td>
<td>2.94</td>
<td>3.69</td>
</tr>
<tr>
<td></td>
<td>0.86</td>
<td>0.87</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Note. Rating Scale: 4 = Mastery, 3 = Satisfactory, 2 = Developing, 1 = Unsatisfactory.

The response rate was 93.0% in 2004–2005, 94.7% in 2005–2006, and 100% in 2006–2007. Supervisor ratings aggregated over 3 years indicated an increase in professional competencies in all six critical domains of school psychological practice: Use of assessment in a problem-solving context, knowledge and skill in consultative problem-solving skills, use of academic intervention strategies, use of behavioral intervention strategies, use of data to monitor progress, and competency in conducting professional development activities (See Table 1).

What Was the Impact of Interns’ Practices on Students Served—Number of Students Served Across Tiers

Number of Students Served. Interns provided data regarding the number of students they served within a three-tiered model of service delivery across three school years. Collectively, the interns provided Tier 1 (universal) practices to 87,068 students, Tier 2 supplemental interventions to 9,974 students, and Tier 3 individualized interventions to 17,293 students in Ohio over the course of three school years. Trends in the average number of Ohio’s students served by school psychology interns over these 3 years suggested that interns are serving an increasing number of students at Tier 1 and slightly fewer students at Tier 3 (See Figure 1). This finding is consistent with the intended design of the three-tiered approach to intervention, with its emphasis on prevention and schoolwide services, thus reducing the need for more intensive intervention.

Demographics of Students Served. School psychology interns provided data regarding the race/ethnicity, gender, LEP status, disability status, and economic disadvantaged status of the students to whom they provided Tier 2 (supplemental/targeted interventions) and/or Tier 3 (intensive/individual interventions) services during the 2004–2005, 2005–2006, and 2006–2007 school years. The results are presented in Table 2. The demographic characteristics of the students served by interns at Tier 2 and Tier 3 were representative of Ohio’s at-risk student population.

Impact on Student Outcomes in Terms of Goal Attainment and Aggregation of Graphed Data. The process for measuring impact based on the aggregation of single case designs was piloted by three volunteer universities in 2004–2005, based on an accountability model developed at one
of these universities (Barnett et al., 1999). Five universities participated in providing this impact evaluation data in 2005–2006. By 2006–2007, it was expected that all nine universities would participate in the collection of impact data, although obtaining outcomes for Tier 1 academic interventions was a challenge for some training programs based on differing perceptions regarding the expected involvement of school psychologists in collaboration to design and monitor Tier 1 academic activities.

GAS outcomes were gathered for 518 Tier 2 and 3 interventions provided to Ohio students by school psychology interns in all nine programs during the 2006–2007 school year. These interventions included 253 academic interventions, 254 behavior interventions, and 11 interventions that targeted a combined academic and behavior concern. Examples of academic interventions included classwide choral reading intervention (Tier 1), small-group writing skills intervention with

![Figure 1](image-url)

**Figure 1.** Three-year trend in the average number of students served by interns at each tier.

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>Count</th>
<th>Percentage of Students Served</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>6,494</td>
<td>24.0%</td>
</tr>
<tr>
<td>American Indian or Native Alaskan</td>
<td>13</td>
<td>&lt;0.1%</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>250</td>
<td>0.9%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>797</td>
<td>3.0%</td>
</tr>
<tr>
<td>Multiracial</td>
<td>870</td>
<td>3.2%</td>
</tr>
<tr>
<td>White</td>
<td>18,586</td>
<td>68.8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demographic Categories</th>
<th>Count</th>
<th>Percentage of Students Served</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11,169</td>
<td>41.4%</td>
</tr>
<tr>
<td>Male</td>
<td>15,841</td>
<td>58.6%</td>
</tr>
<tr>
<td>LEP</td>
<td>517</td>
<td>1.9%</td>
</tr>
<tr>
<td>Students with Disabilities</td>
<td>11,037</td>
<td>40.9%</td>
</tr>
<tr>
<td>Economically Disadvantaged</td>
<td>10,147</td>
<td>37.6%</td>
</tr>
</tbody>
</table>
self-graphing (Tier 2), and extra math opportunities to respond with positive reinforcement (Tier 3). The GAS outcomes for academic interventions are presented in Figure 2. For a smaller number of interventions A-B or related designs were used for evaluation allowing for PND and ES summaries. For Tier 1 academic interventions, the mean PND was 85.1% ($SD = 24.6$, $n = 13$ interventions) with a mean ES of $>3$ ($SD = 0.8$, $n = 5$ interventions). For Tier 2 academic interventions, the mean PND was 77.8% ($SD = 24.6$, $n = 72$ interventions) with a mean ES of 2.8 ($SD = 33.3$, $n = 37$ interventions). For Tier 3 academic interventions, the mean PND was 73.8% ($SD = 26.1$, $n = 66$ interventions) with a mean ES of 2.4 ($SD = 3.16$, $n = 20$ interventions).

The GAS outcomes for behavior interventions appear in Figure 3. Examples of behavioral interventions included classwide “Catch ‘em Being Good” intervention for student engagement

![Figure 2](image1.png)

**FIGURE 2.** GAS outcomes for academic interventions by level of intervention.

![Figure 3](image2.png)

**FIGURE 3.** GAS outcomes for behavioral interventions by level of intervention.
(Tier 1), small-group social skills intervention (Tier 2), and behavioral contracting (Tier 3). For Tier 1 behavior interventions, the PND was 59.3% ($SD = 42.1, n = 21$ interventions) with a mean ES of 2.7 ($SD = 3.0, n = 11$ interventions). For Tier 2 behavior interventions, the mean PND was 65.0% ($SD = 27.3, n = 38$ interventions) with a mean ES of 0.9 ($SD = .94, n = 12$ interventions). For Tier 3 behavior interventions, the mean PND was 78.5% ($SD = 28.4, n = 31$ interventions) with a mean ES of 2.9 ($SD = 3.1, n = 15$ interventions).

**DISCUSSION**

The accountability climate reflected in federal initiatives (NCLB, program accreditation) and emphases on scientifically based and empirically based instruction and intervention continue to have an impact on the professional practice of educators and school-based related service providers (Cowan & Skalski, 2008). The mandate for educational accountability, according to Berry and Eddy (2008), has transformed how professional practices are evaluated by prioritizing a summative evaluation to demonstrate effectiveness over formative evaluation. Professional standards for training and service delivery also emphasize accountability and demonstrating the positive impact of training and services. In this study, a model for evaluating the professional practice of school psychology interns across nine Ohio university training programs was described. The results of the evaluation provide evidence of the skill development of the interns over the course of the school year in six key domains of school psychological practice. The direct impact of the internship experience on the professional development of interns is critical at this point in time with the requirement for greater sophistication in decision making needed for systemic, multitiered approaches to service delivery. This evaluative study expands previous research (e.g., Goodman & Young, 2006) by including a description of the impact that school psychology interns have on the students they serve. Descriptive measures (i.e., the number of students served by tier and by demographic characteristic) and outcome measures (i.e., GAS, ES, and PND) provide important, meaningful evidence of the positive impact of school psychology interns on the students and school communities they serve. Taken together, this model for evaluating a school psychology internship program statewide depicts outcomes that are readily understood and meaningful to key stakeholders whose support is needed to maintain and continuously improve the internship program. In addition, formative evaluation priorities are given nearly equal weight in this model for a statewide evaluation of a school psychology internship program, and the underlying methods used by interns emphasize formative evaluation (e.g., practices used in RTI).

Although all nine university training programs have NASP approval, variations in the specializations of the faculty, program themes, and quality of the graduate student applicants have generated distinctions among the programs over the years. The statewide evaluation model is the product of explicit discussion and consensus regarding the six critical domains of school psychology practice that future school psychologists need to have mastered by the end of their internship. These training priorities, along with standardization across all nine university programs regarding the methods for synthesizing academic and behavioral outcome data across a multitiered approach to service delivery, have required that university faculty revisit how they prepare their graduate students to complete the requirements of internship and the internship evaluation successfully. Thus, all universities have engaged in data-based decision making on their individual and collective training practices, using outcome data to improve training practices, consistent with the goal of NASP performance-based evaluation. Through this data-based decision making, university training programs continue to examine their curriculum and field experiences with a focus on improving outcomes prioritized in the evaluation model. Evidence that this is occurring can be found in the 3-year trends depicting an increasing number of students served at Tier 1 and the decreasing number of students served at Tier 2, suggesting a change in training preparation to focus on universal prevention and early intervention activities as priorities.
Limitations of the Evaluation Design and Procedures

The evaluation derived from the functional purpose of providing outcome evaluation data to the ODE; as such, the evaluation was guided by this goal and not primarily by research goals. Thus, there are some inherent limitations from a pure evaluation and research perspective. The reliance on descriptive outcome measures that are based on interns’ self-report, a necessity given the constraints of collecting data statewide without a budget to support independent data gathering, is the first limitation of this evaluation. Reporting the number of students and their demographic characteristics was dependent on interns’ accurate record keeping and understanding of the three-tiered model. This limitation was addressed, in part, by requiring internship supervisors to approve the Demographic Data Collection Sheet.

A second possible limitation was that the levels established in the GAS process were determined by the intern and may be biased to produce artificially positive outcomes (in the case of goals that are too modest) or artificially negative outcomes (in the case of goals that are too ambitious). Although interns’ judgments regarding the GAS outcomes of the students they served were conducted under the supervision of the field-based and university internship supervisors, no evidence was gathered statewide regarding the accuracy and reliability of the interns’ judgments of the goal attainment. Future evaluations should require an independent review by a second rater and the reporting of inter-rater agreement for each intervention included in the statewide evaluation. Previous studies using these evaluation methods for an individual program used extensive permanent products submitted by trainees (e.g., technical adequacy checks, graphs, intervention scripts). These procedures may be feasible at the individual university level, but were difficult to include in a statewide evaluation that required quick action, efficiency, and collaboration among a diverse group of university programs.

A third limitation of this evaluation model is the lack of a single instrument for assessing intern competencies for use across all of the university training programs. Although the evaluation model was designed to promote a cohesive model for accountability among all of the university training programs while still preserving training programs’ right to self-determination, future research should use a single instrument by all of the participating universities. Where interns are supervised by more than one field-based internship supervisor, independent ratings from both supervisors should be gathered and analyzed to calculate inter-rater agreements as a measure of reliability.

Another important limitation of this evaluation is that interns had the choice of submitting their most successful interventions. Thus, the interventions for which the outcomes were reported should be considered as exemplars of their service delivery efforts and potential quality of interns’ services to students but not necessarily representative of their work in general. Purposive sampling has many important uses in school intervention evaluation (e.g., Bohanon et al., 2006). On the plus side of purposive sampling of exemplars, interns have a say in the meaningfulness of cases and outcomes that are analyzed. This may be important in that intervention outcomes are often out of the complete control of interns. Other options include random selection from the total pool of interventions planned and implemented during the internship year. Analysis of options for data selection and inferences from the data would be addressed carefully by state agencies in evaluation design and reporting.

Other limitations include a lack of specific information on evidenced-based interventions that were monitored at district and program levels. A concise state-level method of describing and coding empirically based interventions from program data may be helpful in future evaluation plans (i.e., Horner, Carr, McGee, Odom, & Wolery, 2005; Kratochwill & Shernoff, 2004). Selection of the key outcome measures and analyses may be refined as well. Although individual interns may have included measures of intervention acceptability in their documentation of intervention outcomes, social validity data were not included as outcome measures in this evaluation model. Future research
should consider adding a common social validity measure for use in the evaluation of the internship program. Last, use of summary statistics such as ES as well as others may be improved (e.g., Marquis et al., 2000; Olive & Smith, 2005; Thompson, 2007).

A final limitation was the lack of more specific information about the field-based internship supervisors who contributed data to this evaluation. Future research should treat the field-based supervisors as participants in the study and, consequently, gather pertinent information regarding their training, professional experiences, years of service, and demographic characteristics.

**Conclusions: Summative, Formative, Doable, and Comprehensible**

This article outlined an evaluation design that was cooperatively negotiated across all state internship programs. A goal of the internship evaluation model was to develop methods that were feasible and sustainable resulting in useful data for maintaining state internship support as well as enabling timely data-based adjustments to training statewide as needed. Formatively, the evaluation served to highlight critical national as well as state initiatives and training directions emphasizing intervention services and data-based decision making. Summatively, the methods and data show the extent and potential quality of interns’ services to students. This evaluation report extends the professional literature on professional training and practice and program accountability by presenting a model, applied statewide, to meet NASP training standards for evaluation of impact of trainees’ practices, performance-based assessment, and evaluation for program accountability and improvement. The evaluation builds on program-level efforts for monitoring and reporting internship outcomes. The evaluation methods also could have potential use by school psychologists in demonstrating their impact and effectiveness, as well as by school psychological service units in larger districts or cooperatives that aim to aggregate data to demonstrate impact and adhere to NASP guidelines for professional practice, which encourage school psychologists to demonstrate the positive impact of their services.

In the future, the internship evaluation model will need to weigh alternatives for increasing accuracy checks and evaluating the representativeness of data in light of the need for utility and sustainability of procedures. The evaluation will continue to evolve, but at the least, the data show the comprehensiveness and direction of training during internship, and the potential for interns’ contributions to improved student outcomes. Future research should explore the generalizability of this model to internships and school psychological service units in other locales.

**Appendix**

**Items from University Training Program’s Internship Competencies Checklists Used to Measure the Six Critical Domains**

*Use of Assessment in a Problem-Solving Context*

- Understanding and using assessment in a problem-solving context:
  - (a) Use data to demonstrate student problems/needs
  - (b) Use data to demonstrate student outcomes
    
    *Four universities use this same item*

- [Under the subheading “Problem Solving”] Uses empirical intervention-based strategies to clarify and identify the problem and target behaviors

- [Under the NASP domain “Data-Based Decision Making & Accountability”] Demonstrates fluency in the problem-solving process
- Uses an ecological/problem-solving model
- Demonstrates knowledge of a systematic problem-solving process
- Analyzes the problem-solving processes being used

Consultative Problem-Solving Skills

- Displays knowledge and skill in consultative problem solving
  (a) Models support for problem-solving initiatives at individual, school, and system levels
  (b) Applies a complete and systematic problem-solving process that includes: identification and clarification of problem situation; analysis of factors related to the problem, implementation and monitoring of interventions, and evaluation

[Four universities use this same item]

- [Under the subheading “Collaboration”] Employs collaborative strategies across all phases of problem solving, from problem identification to problem analysis, intervention planning, implementation, and evaluation
- [Under the NASP domain “Consultation and Collaboration”] Communicates effectively with school personnel
- [Under the subheading “Consultation Skills”] Follows problem-solving steps
- Collaborates effectively with other persons during planning and decision-making process
- Explains and demonstrates collaborative problem solving within the context of a problem-solving process

Academic Intervention Strategies

- Knows when and how to use empirically validated academic intervention strategies
  (a) Knows empirically validated components of effective academic intervention (e.g., immediate feedback, opportunities to respond, contingencies for accuracy)
  (b) Knows empirically validated instructional interventions (e.g., peer-assisted learning, listening previewing, practice strategies)

[Three universities use this same item]

- [Under the subheading “Intervention”] Implements and evaluates academic interventions at Tier 3 (individual student)
- [Under the NASP domain “Effective Instruction and Development of Cognitive/Academic Skills”] Suggests and is able to apply appropriate intervention monitoring methods:
  (a) Understands intervention acceptability as a factor influencing use of interventions
  (b) Supports intervention integrity through development of appropriate monitoring techniques
  (c) Assists in designing and implementing data collection procedures that are appropriate to the nature of the intervention, its goals, and relevant child and environmental factors
- [Under the NASP domain “Effective Instruction and Development of Cognitive/Academic Skills”] Utilizes empirically demonstrated instructional methods/interventions
- Uses evidence-supported interventions
- Develops appropriate academic interventions and goals based on students’ varying abilities/disabilities/needs
- Demonstrates ability to develop and implement instructional/learning interventions
Behavioral Intervention Strategies

- Knows when and how to use empirically validated behavioral intervention strategies:
  (a) Knows empirically validated components of effective behavioral intervention (e.g., cueing, reinforcement, skill-training)
  (b) Knows empirically validated behavioral interventions (e.g., reinforcement plans, self-regulation, problem-solving routines)

[Four universities use this same item]

- [Under the subheading “Intervention”] Implements and evaluates behavioral/social interventions at Tier 3 (individual student).
- Uses ecological and behavioral approaches when developing behavior change programs
- Uses evidence-supported interventions
- Develops appropriate behavioral/affective interventions and goals based on students’ varying abilities/disabilities/needs
- Demonstrates ability to develop and implement behavioral/support interventions for children and youth (preschool through high school)

Use of Data to Monitor Progress

- Suggests and is able to apply appropriate intervention monitoring methods:
  (d) Understands intervention acceptability as a factor influencing use of interventions
  (e) Supports intervention integrity through development of appropriate monitoring techniques

Assists in designing and implementing data collection procedures that are appropriate to the nature of the intervention, its goals, and relevant child and environmental factors

[Four universities use this same item]

- Sets up technically valid procedure for data-based progress monitoring and intervention evaluation
- Uses data to evaluate the outcomes of services
- Uses data to monitor progress
- Uses data to evaluate outcomes
- Evaluates intervention effectiveness

Conducting Professional Development Activities

- Conducts training activities for professional staff and parents/caregivers:
  (a) Assesses potential training needs
  (b) Develops training plan
  (c) Conducts/assists with training, working toward an effective presentation style
  (d) Evaluates training impact/outcomes

[Four universities use this same item]

- Plans, engages in, and evaluates staff development activities (e.g., presentations to staff or parents, ongoing technical assistance)
• Provides effective in-service
• [Under the subheading “Developing and Training Staff, Parents, and Students”] Designs and co-implements staff/parent/student development activities

REFERENCES


Psychology in the Schools DOI: 10.1002/pits


