CONDUCTING AN IN-DEPTH EVALUATION OF A GIFTED PROGRAM

KRISTIE SPEIRS NEUMEISTER, PH.D., AND VIRGINIA H. BURNEY, PH.D.

Learn More at giftED19 Dec 4–6!

PURPOSE OF EVALUATING GIFTED PROGRAMS[1]

Districts that have devoted resources to programs for gifted and talented students hold a vested interest in understanding the outcomes of such programs for the students they serve. Program evaluation, defined as the use of sound research methodology to form an appraisal of the processes and outcomes of a program intended to fulfill a social need (Royse, Thyer, & Padgett, 2016), provides a systematic way of assessing these outcomes for stakeholders.

Texas Administrative Code (19 TAC §89.5) requires that programs be evaluated for effectiveness, and these evaluations can serve multiple purposes. Formative evaluations assess strengths and challenges of the current programming models to determine recommendations for improving efficiency and effectiveness. As a matter of principle, districts also have an ethical responsibility to conduct formative program evaluations to ensure that the programs they have developed for their gifted and talented students are grounded in best practices as outlined in the Texas State Plan for the Education of Gifted/Talented Students (Texas Education Agency, 2019) and therefore are meeting the cognitive, social, and emotional needs of this group of exceptional learners. In the spirit of continual improvement, even districts with strong programs benefit from formative program evaluation by addressing gaps and areas of challenges.

Summative evaluations of gifted programs also offer benefits for districts. Such evaluations may arise out of necessity for accountability as is the case for districts in

[1] Sections of this article are adapted from *Gifted Program Evaluation: A Handbook for Administrators and Coordinators* (2nd ed.), by K. Speirs Neumeister and V. H. Burney, 2019, Waco, TX: Prufrock Press. Copyright 2019 by Prufrock Press. Adapted with permission.

states with mandates such as Texas. In addition, as a matter of practicality, scarcity of resources requires districts to justify spending; consequently, summative program evaluation can provide documentation of the value of the program.

IMPORTANCE OF CONDUCTING AN IN-DEPTH PROGRAM EVALUATION

When determining program effectiveness, district administrators and coordinators of gifted and talented programs may rely on achievement data to document their program's success. Although such data are important to consider, it addresses only one facet of programming: achievement. To thoroughly assess the extent to which the program meets each of the Texas standards for gifted education (Texas Education Agency, 2019), data need to be collected and analyzed that pertain to *all* facets of the gifted program. Such an approach requires an in-depth evaluation that uses multiple data collection methods (e.g., interviews, surveys, document review, observations) and multiple data sources (e.g., teachers, parents, administrators, students, counselors). Although such a process requires the commitment of time and resources, the results of the data analysis provide a thorough evaluation of the program's effectiveness in adhering to its mission and meeting its goals. Without such a process, erroneous conclusions about program effectiveness may be drawn that ultimately impede progress toward program improvement by cloaking problematic program elements.

The purpose of this article is to provide readers with an understanding of what is involved in conducting an in-depth gifted program evaluation and to share tips and suggestions for consideration in planning such an evaluation.

STEPS IN THE PROCESS

CHOOSING WHO WILL CONDUCT THE EVALUATION

The first step for a district to take is to determine who will conduct the evaluation. Evaluations can be conducted internally by a team of district employees or externally by a team of experts in both evaluation methodology and gifted education. Each has strengths and limitations that merit consideration. Internal evaluations are appealing to districts for several reasons. Internal evaluators may have invaluable insight into the program based on frequent opportunities to see the program in action and acute awareness of stakeholder perceptions and the overall reputation of the program. As internal evaluators operate "within the system," they may have easier access to data and more familiarity with the veracity of the program documents, policies, and implementation. Finally, as the internal team is comprised of district employees, the evaluation has minimal financial impact on the district.

Internal evaluations, however, are not without their limitations. Often the district will assign those most familiar with the program to serve on the evaluation team (such as the program coordinator or program teachers). Familiarity with the program may be disadvantageous in this case, as such team members may be too personally entrenched in the program to objectively analyze the data. Moreover, stakeholders may not be willing to share their real perceptions of the program for fear of retribution or hurting the feelings of those directly involved with the program. In addition, conducting a thorough program evaluation requires technical skills both in research methods (including the ability to design a valid and reliable process for collecting and analyzing both quantitative and qualitative data) and written communication for reporting the findings. Districts may be hard pressed to find such a targeted skill set within their faculty. Finally, when an evaluation is conducted internally,

IN THE SPIRIT OF CONTINUAL IMPROVEMENT, EVEN DISTRICTS WITH STRONG PROGRAMS BENEFIT FROM FORMATIVE PROGRAM EVALUATION BY ADDRESSING GAPS AND AREAS OF CHALLENGES.

stakeholders at large may be more resistant to accepting the findings, particularly if significant changes are being proposed. They may question the credentials of the internal team and therefore the validity of the findings and proposed recommendations.

To address these limitations of internal evaluations, districts may consider hiring outside experts in the field of gifted program evaluation. Best practice in gifted education calls for programs to undergo formal evaluation by outside experts approximately every 5 years (Landrum, Callahan, & Shaklee, 2001). Outside experts hold no preconceived notions of the programming and therefore can approach data collection and analyzing with a more objective lens. Likewise, stakeholders may be more willing to share their true perceptions and opinions of the program knowing the evaluators are not employees of the school district. Finally, with formal training in both research methods and gifted

education, the evaluation process may be more robust and the findings more likely to be accepted by stakeholders.

Districts faced with significant budget challenges, however, may not be able to afford outside consultants to conduct such an evaluation. A solution

tate the coordination of data collection but refrain from participating in the data analysis. In this way, the coordinator's experience with the program could be used to benefit the process while preventing the possible perception of influencing the results.

IN CONSIDERING WHAT DATA TO COLLECT, A RULE OF THUMB IS TO INCLUDE AT LEAST THREE DATA POINTS FOR EVERY PROGRAM ELEMENT BEING ASSESSED.

for these districts might be to conduct their own in-house interim program evaluation while planning and budgeting financial resources for an external evaluation by experts in the future.

DETERMINING THE EVALUATION COMMITTEE

If the district decides to conduct the evaluation in house, the next step is to determine the committee to conduct the evaluation. Stakeholders will place more validity in the process if the committee is comprised of members with broad representation from the district. Although teachers and administrators on the committee need to be knowledgeable about the gifted program, too much personal investment in the program (e.g., a key teacher in the gifted program) may prevent the individual from being as objective as is necessary to interpret the data. If the district has a gifted coordinator, this person may be best utilized as an advisor for the committee who could provide guidance on the timeline for the evaluation and facili-

DETERMINING THE SCOPE OF THE EVALUATION

The scope of the evaluation will also need to be determined and clearly defined by the district. The scope may be determined by a specific concern that has arisen in the district, sparking the support for conducting the evaluation. For example, if a district was losing a high percentage of its most able middle school students to a nearby charter school, it may want to focus the evaluation on examining the gifted and talented program at the middle school level to determine how to strengthen it to appeal more to families. The scope may also be determined based on the resources available to conduct the evaluation. Small districts may be able to assess their gifted program and services in its entirety in one holistic evaluation, whereas larger districts with multiple programs and services may need to evaluate particular elements of their programs one at a time each year in order to make the process more manageable for their team.

ORGANIZING AN INTERNAL EVALUATION TIMELINE

Once the scope of the evaluation has been decided, the committee is ready to organize the process. An agreed-upon timeline and a corresponding person responsible for the completion of each step will ensure that the process continues to move at a smooth pace. The timeline should be constructed with careful consideration of other district initiatives and responsibilities. For example, filming of classroom instruction should not be scheduled the week prior to statewide testing or surveys should not be sent home to parents right before winter break, as that time of year tends to be chaotic for families, and therefore response rates may be lower. Districts may also want to build extra time into the process to account for unforeseen interruptions to the data collection process. Building extra time into the timeline will also ensure timely completion, even if certain aspects prove to be more time consuming than anticipated.

DATA COLLECTION

One of the most important tasks of the evaluation committee is collecting all the data needed for analysis. In considering what data to collect, a rule of thumb is to include at least three data points for every program element being assessed. This process, referred to as triangulating the data, strengthens the validity of the findings presented because the conclusions are reached only after considering commonalities noted across the results from each of the three data points, rather than relying too heavily on any one data point in isolation. For example, if the program area being evaluated is curriculum and instruction, the data collection points might include a review of the curriculum maps and other curricular documents, classroom

observations, and survey questions related to curriculum and instruction.

Evaluators also need to understand the purpose of different types of data collected and determine the combination of data that will yield the most comprehensive understanding. Quantitative data, such as achievement test scores, provide solid, objective information or facts about program elements. Quantitative results are needed to provide evidence for justifying program recommendations. Numbers alone, however, do not explain the results they provide. For this reason, we recommend a mixed methods approach that includes both quantitative and qualitative data as it provides the most comprehensive understanding of the program.

TYPES OF DATA

Interviews. Interviews provide a valuable source of data for program evaluations. When conducted well, interviews allow evaluators to hear in the stakeholders' own words their perceptions of program strengths, areas of challenge, and suggestions for improvement. Although the number of interviews needed will depend on the size of the district and scope of the evaluation, a general recommendation is to continue interviewing until data saturation has been reached. Data saturation occurs when the interviews are no longer generating new information or perspectives to consider. Multiple representatives from each stakeholder group should be included in the interview process.

Two types of interview structures are most helpful in evaluations: semi-structured and focus groups. In the semi-structured approach, the interviewer prepares a general list of open-ended questions on all aspects of the program. The interviewer pays close attention to the participants' responses and asks follow-up prompting questions to elucidate a more thor-

ough understanding of the points that appear to be the most important to the interviewee. This process keeps the interview focused, but allows the participants the opportunity to elaborate more on key points as they deem necessary. Another type of interview, the focus group, may also be beneficial in an evaluation. Focus groups typically consist of 6–10 members of a stakeholder group sitting in a circle offering responses and holding a discussion based on the interviewer's questions. Focus groups are appealing to evaluation teams because they allow for more stakeholders to participate without taking excessive amounts of time to complete. Moreover, they provide a "check" on extreme views, as the interviewer can easily see the reaction of other participants to individual responses. Focus groups are not without their limitations, however. Sometimes a "group think" mentality can emerge where one person brings up an issue or concern, and others internalize that as their own even though they may not have endorsed this view had they been asked about it in a one-on-one interview setting. Power dynamics can also affect the quality and types of responses given during focus groups.

Surveys. Survey responses are an invaluable source of data in a program evaluation, as they allow the evaluators access to the perceptions of each program component from all stakeholder groups. Although a full treatment of all factors to consider when designing surveys is beyond the scope of this article[2], the following tips for creating and administering survey questions should assist the evaluation team in getting started.

- **Carefully consider the length of the survey.** Evaluation teams should be cognizant of the overall length of the surveys, including only the most salient questions from each program area, as sur-

veys that take longer than 10 minutes may not be completed.
- **Design surveys for multiple stakeholder groups.** Surveys should be designed for dissemination to multiple stakeholder groups, including administrators, teachers, parents, and students. By asking parallel questions (when appropriate) to all stakeholder groups, evaluators can determine if perceptions are similar or discrepant across groups.
- **Include both forced choice and open-ended questions.** Both forced-choice response questions and open-ended questions are valuable. Forced-choice responses allow for a more direct comparison of responses across stakeholder groups to determine overall differences in perceptions, attitudes, or opinions. Open-ended questions allow for the respondents to write narrative comments, enabling them to elucidate on earlier forced-choice response choices or to introduce additional topics or opinions related to the gifted program that were not previously explored.
- **Consider the wording of questions carefully.** It is always surprising how difficult it is to construct carefully worded questions that are specific, nonbiased, and easily understood by the respondents. The following tips will help evaluation teams in their construction of concise, effective survey questions.
  - *Construct neutral questions.* Avoid including adjectives that may trigger an emotional reaction within the respondent that may cue him or her to respond a certain way. For example, providing parents with statements such as "My child is bored in language arts class" or "My child feels isolated in the gifted program" and asking them to circle

their level of agreement with the statement (*Strongly Agree, Agree, Undecided, Disagree,* or *Strongly Disagree*) may cue them to worry that their child *is* bored or *is* isolated and then negatively bias their response accordingly.

- *Use simple language.* Survey questions should be as clear and concise as possible to minimize confusion. In addition, it is important to keep in mind the different levels of educational experiences and cultural backgrounds of the respondents when phrasing questions.
- *Make sure questions have only one aspect to consider.* Although seemingly obvious, a frequent flaw of survey questions is that they contain multiple items. The responses for these questions cannot easily be interpreted because one cannot be certain which aspect(s) the respondent was referring to in the answer.
- *Balance the rating scale.* Including an equal number of positively and negatively balanced response choices can help avoid inadvertently skewing the results.
- *Field test survey questions before dissemination.* Once a draft of the survey is completed, field testing among a small group of potential respondents can ensure that the questions are clear, no important topics were missed, and the completion time is not too long.
- *Utilize multiple methods of dissemination.* To ensure a greater response rate and more representation among stakeholder groups, the evaluation team will want to consider multiple methods of survey dissemination.
- *Emphasize the importance of survey completion.* District leadership can assist the evaluation team in promoting the importance of survey completion by stakeholder groups.

Observations. Observations of program activities, including classroom instruction, can provide a valuable data source for understanding the experiences of gifted students. Evaluators are encouraged to record as much detail as possible during the observations, as the importance of details noted during the observations may not be readily ascertained until analyzed in concert with data from other sources such as interviews and surveys.

DOCUMENTS

Program documents include everything from fliers describing the program to full program handbooks. Reviewing all program documents at the beginning of the program evaluation is beneficial for gaining a foundational understanding of the program. The documents often raise questions that can be addressed during interviews with stakeholders. Program documents may also provide a window into understanding the intent of the program, as many documents may have been drafted in the beginning planning stages of program development.

Although documents include a wealth of information on the program, evaluators are reminded of the importance of triangulating the information gleaned from the documents with other data sources as well. For example, documents may be excellent data sources for comparing the program's stated goals and policies with actual practice; in other words, does practice match the written description of the program, or is the program on paper only?

TEST SCORES

Scores from tests used in the identification process may be reviewed against the published identification procedures and state guidelines to determine fidelity in the selection of students. Performance on state achievement assessments may be used to determine if identified gifted students are earning scores in the highest performance categories. Finally, as a measure of rigorous curriculum and instruction, scores from computer adaptive, norm-referenced achievement measures may be used to determine if gifted students are reaching their expected growth targets.

PROGRAM ARTIFACTS

Program artifacts can provide a rich source of data for examining program goals that may otherwise be difficult to observe. For example, many gifted programs have mission statements emphasizing the goal of developing skills such as creative thinking, leadership, and citizenship. Documentation of the development of skills such as these may be evidenced in program artifacts, including projects submitted for academic competitions, service learning projects, and presentations to culminate units of study.

INTERPRETATION OF DATA

Collectively, myriad data sources will give evaluators a more comprehensive understanding of the effectiveness of the gifted program. Quantitative data in the form of survey responses and test scores offer evidence of trends, gaps, and differences in perspectives, skill development, and achievement. Accompanying qualitative data may provide interpretations for these results. The results of both types of data in concert will provide a rich understanding of the program's areas of strength as well as areas of challenge, enabling the district to develop a roadmap of recommendations for future improvement.

Although it is imperative to triangulate findings by using more than one data source, all data sources are not equal and should not be given equal weight in the interpretation. For example, if the internal evaluation committee is interested in assessing program effectiveness for academic achievement, it may examine the results from achievement test data as well as survey questions asking students to rate to what extent the pro-

gram has facilitated high achievement. If the survey results report that students feel the program is facilitating high achievement, yet the achievement test scores themselves suggest that the students in the program are not achieving at high levels, the test data should be given more attention than the survey data. Perhaps the students have a different perception of what high achievement means compared to district administrators. This example highlights another purpose of triangulating data; it can help to identify areas that need further investigation, such as why the data indicate a discrepancy between students' perception of high achievement and actual achievement test scores.

One must also pay close attention to how the findings from survey data are used to influence program decisions. Survey data are just perception data; the data give an indication of how program stakeholders perceive various program elements. Although these results may be beneficial to program administrators, it does not mean that the survey respondents are "voting" on what elements the program should include. Those decisions should be made based on best practices and state requirements. Results of surveys, however, may help the gifted coordinator know how to target professional learning and parent information meetings to help clarify any misperceptions or lack of understanding that stakeholders may have about any of the program components.

WRITING THE REPORT

After the data are analyzed, the committee is charged with summarizing the findings and developing an action plan. The evaluation team is encouraged to first present the findings highlighting areas of strength. Stakeholders have invested time and energy into the program, and their attention to the program needs to be recognized and appreciated. Beginning with a recognition of the positive facets of the program will also predispose those in charge of the program to be more receptive to hearing the challenges.

Our recommendation is to use the term *challenges* instead of *weaknesses* to describe those areas of the program that are less well developed. The word *weakness* connotes a flaw, with the indication that the program coordinator should have been able to prevent or fix this flaw. In contrast, the subtle change in semantics through the use of the word *challenges* indicates the recognition that some areas of programming are difficult to design and implement effectively because of a whole host of external factors. To address these areas of challenge, the program coordinator will need district support. Areas of challenge should be perceived as being owned by the whole district rather than perceived as the sole responsibility of the program coordinator to address.

TEXAS ASSOCIATION FOR THE

GIFTED & TALENTED

PARTNERS IN

Gifted Education

JTAYLOR
EDUCATION
DEPTH & COMPLEXITY
Resources & Professional Development
Est. 2006

Riverside
Insights

Pearson

Prufrock
Press Inc.

THINK
LAW

Once the areas of challenge are described with supporting data, the evaluation team should provide a list of recommendations to address each of these areas. We suggest grouping the recommendations into two categories: quick-wins and long-term recommendations. Quick-wins are those recommendations that can be implemented immediately. For example, if the program evaluation indicated that communication with parents is an area of challenge, a "quick win" might be to have each program teacher send a monthly e-mail update to the parents of students in his or her classroom to highlight what has been learned and share pictures of instructional activities and products that reflect the goals of the gifted program. Long-term recommendations, on the other hand, require more time to implement due to their complexity. In this example, a long-term recommendation for improving communication may be to develop a quarterly series of presentations for parents, including an overview of the program in the fall, two presentations on characteristics and social and emotional needs of gifted students during the school year, and a culminating spring showcase of student products or performances reflecting what students had learned in the gifted program throughout the year.

In the interim years between the formal evaluations conducted by outside experts, internal evaluations that follow the process outlined in this article can be of great benefit to the district. The findings and subsequent recommendations gleaned will be invaluable in guiding future program development and improvement, thereby ensuring that the program continues to fulfill its mission of meeting the needs of gifted students.

REFERENCES

Fowler, J. F. (2014). *Survey research methods* (5th ed.). Los Angeles, CA: SAGE.

Landrum, M. S., Callahan, C. M., & Shaklee, B. D. (2001). *Aiming for excellence: Gifted program standards. Annotations to the NAGC Pre-K–Grade 12 Gifted Program Standards.* Waco, TX: Prufrock Press.

Royse, D., Thyer, B., & Padgett, D. (2016). *Program evaluation: An introduction to an evidence-based approach* (6th ed.). Boston, MA: Cengage Learning.

Texas Education Agency. (2019). *Texas state plan for the education of gifted/talented students.* Austin, TX: Author.

Texas Administrative Code, Title 19, Part II. Retrieved from http://ritter.tea.state.tx.us/rules/tac

Kristie Speirs Neumeister, Ph.D., is a professor of educational psychology at Ball State University where she directs the gifted licensure program and teaches graduate courses in educational psychology relating to gifted education. Throughout her career she has focused her research on aspects of perfectionism and how perfectionism influences the achievement of academically gifted students. Dr. Speirs Neumeister was the recipient of the National Association for Gifted Children's *Gifted Child Quarterly* Paper of the Year and Early Scholar Awards. She is also the coauthor of *An Introduction to Gifted Education: The Complete Kit for Facilitators, Coordinators, and In-Service Training Professionals* and *Gifted Program Evaluation: A Handbook for Administrators and Coordinators* (2nd ed.).

Virginia H. Burney, Ph.D., is on the educational psychology faculty at Ball State University and is a consultant to school districts in the area of high-ability education. She has worked K–12 as a principal, school counselor, guidance director, math teacher, and school board member. Ginny is a past president of the Indiana Association for the Gifted, served 5 years on the Board of the National Association for Gifted Children, and is the 2012 NAGC David A. Belin Advocacy Award winner. She is the coauthor of *An Introduction to Gifted Education: The Complete Kit for Facilitators, Coordinators, and In-Service Training Professionals* and *Gifted Program Evaluation: A Handbook for Administrators and Coordinators* (2nd ed.).

Stay Connected! SPRING 2020 EVENTS

TAGT ADVANCE YOUR GIFTED PROGRAM LEADERSHIP CONFERENCE

March 29–31, 2020, Sugarland, TX

An event for G/T administrators, coordinators, and specialists, focusing on best practices and promising programs.

GIFTED + EQUITY SERVING the UNDERSERVED
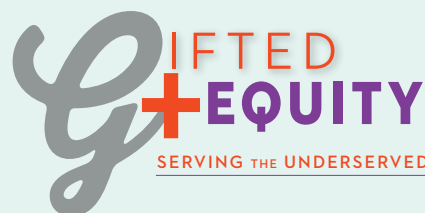
June 8–10, 2020, Denton, TX

Geared toward the G/T leaders and educators who work to ensure equitable access for all gifted students including underserved populations, ELL, and twice-exceptional.