How to *Evaluate*
Your School or District G/T Program

Corina R. Kaul, Brenda K. Davis,
and Susan K. Johnsen, Ph.D.

Colleen is the new director of gifted and talented programs and advanced academic services at Bright Day Independent School District (BDISD). BDISD includes approximately 12,000 students on the outskirts of a sprawling metropolitan area. Ten years ago, the population of pre-K–12 students was comprised of 88% White, 8% Hispanic, 1% Asian, and 3% Black. The demographics of the school district has become increasingly diverse and is now 61% White, 28% Hispanic, 2% Asian, and 9% Black. Currently, the district has identified 11% of the students as receiving special education services, 32% as receiving free and reduced lunch, 15% as English language learners, and 11% as gifted and talented. In addition to Colleen, the Advanced Academics office oversees administration of the district's gifted education program and employs three primary and two secondary G/T specialists. Due to budget cuts, the district has recently changed its elementary pull-out service model to a cluster model where gifted students are grouped together within general education classrooms. At the middle school and high school levels, students are served using advanced courses, mentoring, and enrichment through the arts.

Although Colleen is aware that the effectiveness of G/T services needs to be evaluated annually, the school district has not reviewed the gifted and talented program in more than 5 years. The school board, district-level administrators, and Colleen's Advanced Academics team are all concerned that the gifted program may not be meeting the needs of the gifted student population and that the campuses may not be in compliance with the Texas State Plan for the Education of Gifted/Talented Students (Texas Education Agency [TEA], 2009). Because she is new to the district, Colleen, with members of the Advanced Academics staff, set up informal meetings with members of different constituent groups that include parents, teachers, members of the school board, and administrators to gain their perspectives regarding issues related to the program.

From her conversations, she learns some parents are concerned that minority students are not represented within the identified population of students in

the gifted program. Other parents are concerned that their gifted children are not challenged in the elementary classrooms and attribute the lack of challenge to the cluster model that the district transitioned to a few years ago. Administrators are concerned about disparities between elementary schools because some schools have more students identified than others and few students from traditionally underrepresented student groups have been identified as gifted and talented. Some of the elementary classroom teachers are overwhelmed with the responsibility of meeting the strengths

Hertberg-Davis (2013), the evaluation of programs for gifted students "is necessary in creating and maintaining an exemplary, defensible gifted program" (p. 6). Texas requires school districts to conduct ongoing evaluations of professional development for G/T education (4.4C) and conduct annual evaluations of the effectiveness of G/T services (5.3C; TEA, 2009).

Although the literature reports a few G/T program evaluation models, widespread use or acceptance of any one model is limited (McIntosh, 2015). McIntosh (2015) suggested

(see "A Framework for Program Evaluation," http://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/framework-for-evaluation/main).

To incorporate best practices from Milstein et al.'s (2017) model and other evaluation experts, we propose a Formative Evaluation of Gifted Programs Model (FEGPM) that incorporates stakeholder input and the implementation of evaluation findings (see Figure 1). Four key questions need to be considered during each stage of the evaluation: Are our actions *purposeful* and linked to desired outcomes? Are these actions *useful*? Are these steps *practical*? Will this lead to *accurate* results and reporting?

> ## According to Callahan and Hertberg-Davis (2013), the evaluation of programs for gifted students "is necessary in creating and maintaining an exemplary, defensible gifted program" (p. 6).

and needs of their students who range from identified special education students to those identified as gifted and talented. Members of the school board are sensitive to demographic changes in the district and want the Advanced Academics team to ensure that the district is meeting state requirements. The school board wants Colleen to develop a 5-year plan to ensure that all students who need gifted education services are receiving them.

Colleen wants the evaluation process to identify program strengths as well as gaps in services. With data from the evaluation, she and the Advanced Academics team anticipate that necessary refinements and internal adjustments can be made to embark on a transformational process that will address the needs of the district's gifted population.

PROGRAM EVALUATION MODELS

According to Callahan and

using a content analysis to examine preexisting written documents for qualitative program components and an outcome-oriented evaluation for qualitative outcomes that can be measured numerically. A content analysis includes these components: definition and areas served, philosophy, G/T identification criteria and procedure, goals and objectives of the G/T program, G/T student goals and objectives, curriculum used in the G/T program, personnel and professional development, budget, and program evaluation plans. A quantitative analysis of outcomes may include the effectiveness of G/T identification criteria and procedures, G/T curriculum, and implementation of G/T students' goals and objectives.

Milstein, Wetterhall, and the CDC Evaluation Working Group's (2017) Framework for Program Evaluation proposed four standards for a "good evaluation: utility, feasibility, propriety, and accuracy"

FORM A STAKEHOLDER EVALUATION TEAM

Forming a stakeholder evaluation team gives all vested audiences the opportunity to voice their concerns through their representatives by suggesting questions or giving opinions. Program evaluation should not be conducted in isolation but should include individuals from each of the stakeholder groups: those who operate the program, those who benefit from or receive the program services, and those who have the authority to decide and implement any potential recommendations (Milstein et al., 2017). For example, one district formed a Gifted Education Study Group that included parents, administrators, and K–12 teachers who determined the direction of the evaluation (Bohn, 2000). Keeping in mind the Formative Evaluation of Gifted Programs Model, the selection of the stakeholder evaluation team needs to be *purposeful* (i.e., think politically to get members on the team who are influential, have authority, and are representative of different groups), *useful* (i.e., select individuals who have a vested interest), *practical* (i.e., assemble a group neither too large nor too small who will have opportunities to contribute to the evaluation), and consider *accuracy* (i.e.,
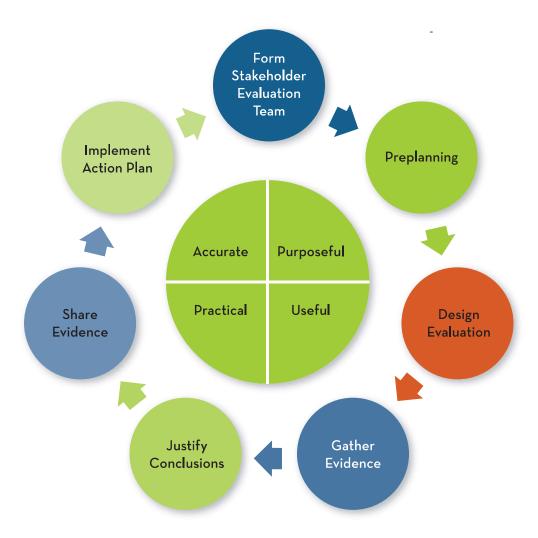
FIGURE 1. Formative Evaluation of Gifted Programs Model (FEGPM).

select members who have experiences or backgrounds in gifted education).

In determining specific stakeholder evaluation committee members, district representatives could consider actions that will enhance representativeness and constituent buy-in such as approaching a G/T parent group for parent nominations or asking educators for teacher nominations. According to the Texas State Plan, educators involved in making decisions for the program should have training in the nature and needs of gifted students (TEA, 2009). Accordingly, gifted education teachers, administrators, and gifted coordinators who meet these criteria

and have a vested interest in improving the gifted program could be included.

As Colleen is new to the district and because she wants to get buy-in from the various constituent groups, she decides that she needs to get recommendations for stakeholder evaluation team members who are influential and will be actively involved. She asks the district G/T parent support group to nominate two parents for the committee. Because many of the concerns are from elementary parents, teachers, and administrators, the district superintendent nominates two principals from different elementary schools across the district (representing one school with a strong G/T program and one school needing improvement). Each of the remaining principals nominates an elementary teacher who teaches gifted

students, and teachers from those schools vote on the nominated teachers to select two elementary teacher representatives. A school board member with a vested interest in gifted education, one G/T curriculum specialist, and Colleen complete the stakeholder evaluation team.

PREPLANNING

The primary objective during the preplanning phase of the evaluation process is to identify the *purpose* of the evaluation. Depending upon the purpose and the questions asked, evaluations can be either formative or summative (Moon, 2012). Summative evaluations are typically conducted by individuals outside of the organization and emphasize the effectiveness of the program and stu-

dent outcomes (Gallagher, 2006). In contrast, formative evaluations focus on answering questions that monitor progress toward a goal, such as "Is this program functioning like it is described in the program document?" or "How can the program be improved?" Schools use formative evaluations for their own purposes to determine program effectiveness (Gallagher, 2006). Accordingly, most gifted program evaluations fall under the category of formative evaluations.

As part of addressing *purpose*, the stakeholder evaluation team needs to consider *why* it is evaluating the program and *who* or *what* is the impetus for the evaluation. For example, is the primary purpose(s) of the evaluation to improve the G/T program, analyze how to increase diversity in the gifted program, select G/T curriculum, examine AP student outcomes, and/or to reduce services and personnel to cut costs? Potential reasons for the evaluation may be identified by reviewing the main complaints of the various parties and considering the viewpoints or goals of persons or groups raising the issues. Reviewing constituent concerns may provide potential evaluation questions and help clarify its purpose.

Second, the evaluation needs to be *useful*. After the evaluation, who will receive the results, how will the results be used, and who will use the evaluation results? This early planning increases the probability of the eventual use of evaluation results (Fleischer & Christie, 2009). The team should consider stakeholders' agendas because a consideration of the concerns, needs, and viewpoints of the intended audience(s) will help to increase the likelihood that the evaluation will address different constituent groups' issues and will improve programs and services for gifted and talented students.

Third, *practicality* should be considered. Practical considerations include the evaluation budget and timeframe. Budgets may need to include money allocated for instru-

ment development, observations, data analysis, reports, and external support. An evaluation in which the entire allocated budget is spent on data collection without consideration for data analysis is ultimately of very little value. Other practical items to keep in mind include the feasibility, accessibility of people involved, the roles/involvement of different stake-

[A]ccuracy in data collection, analysis, and reporting is essential.

holder team members, and/or availability of instruments for proposed data collection.

Finally, *accuracy* in data collection, analysis, and reporting is essential. For example, if data are going to be collected across schools and classrooms, then a specific procedure will need to be followed and instruments need to be carefully designed for validity and reliability purposes.

After hearing from all sides, Colleen and the Advanced Academics team brief the evaluation team on concerns relayed by the constituent groups shared during introductory informal meetings. The evaluation team members remind themselves of the common concerns of constituent groups such as school boards, superintendents, principals, G/T coordinators, and G/T teachers (TEA, 2015). This helps keep the political implications of the evaluation in the forefront of their minds. The school board wants to know if the gifted program is meeting state requirements, the administrators want equity across all programs, parents want access to challenging programs, and the teachers want more support for gifted students

in the classroom. Given these concerns, the team decides that a formative assessment will be needed to determine if the district is in compliance with the Texas State Plan with respect to equity of access, program services, and implementation across schools. Furthermore, if the district leaders want to go above and beyond the compulsory compliance level and make changes to examine the effectiveness of the program for gifted students, they believe that baseline data on student outcomes need to be collected as well. The team works together to draft some initial evaluation questions and suggest potential methods for evaluating these questions.

DESIGN EVALUATION

When designing the method for evaluation, an evaluation team will want to revisit the key points (purposeful, useful, practical, accurate) in the preplanning process as they determine the direction of the evaluation, specific evaluation questions, specific procedures, data collection, and the time frame for collection. They will also need to decide if the evaluation will be conducted internally or if they will hire an external evaluator. Depending upon the expertise within the school district and the political climate, an outside evaluation may be important to design or select data-gathering instruments that limit bias to the evaluation. Keeping the evaluation questions, evaluation sources, and evaluation methods *purposefully* linked to main constituent concerns is important throughout this stage.

Evaluation questions. The *purpose* of the evaluation should drive the overarching evaluation questions. The evaluation team should revisit perceived problems and potential evaluation questions to finalize the main evaluation questions. Callahan (2004) emphasized that these questions should be relevant (i.e., address the function of the specific G/T program), useful (i.e., helpful in making program decisions), and important (i.e., provide

data that address impactful components and outcomes). Evaluation questions prioritize stakeholder issues and those relating to the central functions of the G/T program.

Evaluation questions will have a direct influence on the type of information gathered, the source of the information, how it will be collected (e.g., the instruments that would measure those elements), when it will be collected, and how the information will be used. To maximize *usefulness*, information should be gathered from individuals with different perspectives (e.g., parents, teachers, and administrators). Consideration of evidence that constituents will find credible should be considered (Centers for Disease Control and Prevention [CDC], 2011). Data collection from multiple sources, or triangulation, also enhances the validity, or *accuracy,* of the research (VanTassel-Baska, 2006).

Most likely, Texas school districts should investigate the compliance of their G/T programs to the Texas State Plan (TEA, 2009). An example of an evaluation question addressing this topic might be written, "Is Texas School ISD's G/T program in compliance with the Texas State Plan?" Other evaluation questions might address stakeholder perceptions, specific program activities, outputs, or resources. Questions that address results might include, "What skills learned in the district's summer professional development for G/T teachers do teachers report implementing in the past year?", "In what ways have educators and parents/community formed collaborative partnerships for G/T students during the past year?", or "In what programs did G/T students' products receive district level or higher recognition?"

Evaluation method/data sources. Depending on the purpose(s) of the evaluation, instruments are designed to measure facts, goals, beliefs, processes, or outcomes and can be collected through observations, tests, surveys, document reviews, or focus groups (VanTassel-Baska, 2006). There are advantages and disadvantages to the various methods, such as document review, observation, survey, interview, and focus group, which should be considered when selecting the evaluation method. Most frequently, perceptions of constituents involved in a gifted program are gathered using a survey or questionnaire (Bohn, 2000; Callahan, Moon, & Oh, 2017; Walker & VanderPloeg, 2015) or interviews (Long, Barnett, & Rogers, 2015; Swanson, 2016). Checklists, such as those developed and piloted by Matthews and Shaunessy (2010) for use in assessing compliance with NAGC standards, may also be used for gifted program evaluation. Open-ended questionnaires are better for smaller numbers of respondents. Other measures that can establish a baseline and investigate growth include state-level standardized test scores (Olszewski-Kubilius, Steenbergen-Hu, Thomson, & Rosen, 2017), national standardized tests (Warne, 2014),

> **. . . it is important to examine the validity of the instrument and the degree to which the instrument accurately measures the evaluation questions.**

and AP enrollment and test scores (Gagnon & Mattingly, 2016). In her review of 20 district evaluations, VanTassel-Baska (2006) reported that quantitative data were collected using a parent survey, an educator survey, and classroom observations, and qualitative data were collected using individual interviews, structured focus groups, and document review.

Evaluation instruments. *Practical* considerations related to selecting or designing instruments include assessing alignment to the evaluation questions and the fit within the budgetary constraints and timelines. The evaluation team needs to consider if the district has personnel who have the time and expertise for selecting or designing instruments and analyzing and reporting data from these instruments. This is especially important because many G/T program evaluations that occur are hindered by insufficient training, time, resources, and the difficulty in selecting appropriate evaluation tools and methodology (Cotabish & Robinson, 2012; VanTassel-Baska, 2006). If specific instruments are needed, it is often more expedient and cost-efficient to use a preexisting instrument. However, it is important to examine the validity of the instrument and the degree to which the instrument accurately measures the evaluation questions.

The benefits of developing instruments are the ability to tailor questions to evaluation goals, intended use, and a specific audience. Special care, however, must be taken if the team opts to develop its own instrument. Survey information can be gathered by mail, paper at an event, or using online platforms. Advantages of online surveys compared to traditional administrations include decreased cost, increased access to all participants, and time efficiency because data are automatically gathered and ready for analysis; disadvantages include access and sampling issues (Wright, 2005).

Specification of data analysis is necessary during this stage because it impacts the resources, skills, and data analysis instruments that need to be accounted for in the evaluation plan.

Keeping in mind the budget, personnel, and timeline constraints, the evaluation team designs its evaluation plan based on the evaluation questions it developed to address the concerns of its constituents (see Table 1). Given

TABLE 1
Bright Day ISD Evaluation Plan

Possible Problem Areas	Evaluation Questions	Sources of Information	Assessment Instrument	Data Analysis	Expected Outcome
1. Compliance with Texas State Plan (Sections 1 and 3)	1. To what extent is Bright Day ISD's G/T program in compliance with the Texas State Plan (Sections 1 and 3)?	Bright Day ISD G/T written documents Texas State Plan Administrator, teacher, parent, and student surveys	TEA (2013) Guiding Questions Survey	Content analysis of the Bright Day G/T program using documents and state plan Statistics to describe participants' perceptions of compliance	A summary report comparing BDISD data with TEA compliance indicators to determine if BDISD's gifted and talented program is in compliance with Sections 1 and 3 of the Texas State Plan
2. Equity in access to G/T programs • underrepresentation of minority students • perceived within-district disparities • (see 1.5.1C, 1.5.2C, 1.5.3C, 1.5.4C, 1.6C in Texas State Plan)	2a. To what extent do nominated and identified G/T students reflect the district population?	District and campus demographics (e.g., race, gender, ELL, disability, low-income); District demographics of nominated G/T population and identified G/T population	Demographic forms	Statistics to describe demographics of district and G/T program	A summary report comparing demographics of G/T nominated and identified student demographics for BDISD overall and for each campus
	2b. To what extent do nominated and identified G/T students reflect the individual campus populations?	Students nominated by classroom/grade level and total number of nominations per school	Demographic forms	Statistics to describe demographics of students nominated and identified at each campus/grade level	
	2c. How are instruments used on each campus during the identification process?	Campus use of instruments during nomination and during identification phases	Forms and assessment instruments used during nomination and identification phases	Document review of instruments used at the nomination and identification phases	A description of instruments used in the nomination and identification phases by campus
3. Concerns with lack of curricular and instructional challenge (e.g., depth, complexity, pacing) within the school district and within campuses (see 3.1C, 3.2C, 3.3C in Texas State Plan)	3a. To what extent do teachers, parents, and students perceive a lack of challenge in G/T classrooms?	Teacher, parent, student surveys	Survey	Statistics to describe participants' perceptions of challenge in G/T classrooms	A summary report describing perceptions of challenge in G/T classrooms by campus
	3b. To what extent does the program meet curricular and instructional standards for G/T programs?	G/T teachers' classroom curriculum, including lesson plans, differentiation resources, and learning progressions (above-level work)	Curriculum data form aligned to TEA compliance standards (e.g., depth, complexity, pacing)	Content analysis of curriculum within and across campuses using curriculum documents and state plan	A report of curriculum used by campus and its relationship to TEA compliance standards
	3c. To what extent are teachers implementing differentiation in their classrooms? 3d. How are G/T students clustered?	Classroom observation	Classroom observation instrument (see instrument options in Johnsen, 2017)	Descriptive statistics of observations	A report of the differentiation of the curriculum in G/T classrooms and how G/T students are clustered and interacting with the curriculum
	3e. What are baseline levels for gifted students' performance on the TPSP?	G/T student performance on TPSP	TPSP G/T student rubrics	Descriptive statistics of G/T student performance on rubrics	A summary report describing baseline levels of G/T students' performance on TPSP by campus

the purpose of the evaluation is to collect baseline data, the team decides to collect data on three major areas that include compliance, equity in access, and curricular and instructional challenge. With respect to state compliance, the team will review documents and complete a content analysis to ascertain the level of compliance in the two sections of the Texas State Plan most closely aligned to the major problem areas identified by the constituent groups: Section 1 (Assessment) and Section 3 (Curriculum and Instruction). The team will create surveys for administrators, parents, teachers, and students to gain perceptions regarding the assessment process and curriculum and instruction of G/T students. To answer the equity in access questions, the evaluation team will use demographic data and a document review to obtain needed data. To answer questions pertaining to curricular and instructional challenge, the team will create a curriculum data form aligned to the Texas State Plan and best practices, classroom observations, and surveys with parents, teachers, and students. Principals and G/T specialists will use an observation scale to assess the level of differentiation and teaching practices in the classroom. Baseline levels on TPSP scores will also be collected. Because most of the constituents' concerns stem from elementary campuses, the team decides to focus its evaluation and the observational component on those campuses this year.

GATHER EVIDENCE

Gathering credible evidence is the cornerstone of a good evaluation and is essential in increasing the *accuracy* of the results. In the next phase of program evaluation, information is gathered using the instruments and methods selected in the evaluation design phase. Collection procedures should enhance reliability and validity of the evaluation. If a district does not have skilled personnel in data collection and analysis, training or outsourcing may be required. If using a survey, the questionnaire

should be pilot tested before administration. If information is being collected by observation, interview, or focus groups, it is essential that data collectors be trained in the use of the data collection protocol. Discussion between data collectors after data collection assists in clarifying questions and increasing interrater reliability; cross-checking by other data collectors reduces researcher bias and improves *accuracy* (VanTassel-Baska, 2006). As much as possible, data collection procedures should be standardized. Respondent confidentiality should also be protected. The inclusion of stakeholders and transparency in the process enhances credibility (CDC, 2011).

Gathering credible evidence is the cornerstone of a good evaluation and is essential in increasing the accuracy of the results.

To gather evidence to assess the compliance of G/T program policies with the Texas State Plan (TEA, 2009), a content analysis should be performed where educators compare important terms and vocabulary across documents. To assess written policy compliance, district policies and communication related to G/T services should be collected and systematically analyzed relative to the Texas State Plan. Guiding Questions for Program Review (TEA, 2013) may also be helpful in guiding this process. To examine perceived compliance with the plan, surveys can be designed that include similar questions for administrators, parents, teachers, and students. Recall that content analysis is also appropriate to examine the philosophy of the G/T program, areas of giftedness identified and served, identification procedures and program goals (Carter & Hamilton, 1985, 2004).

Colleen and the evaluation team feel comfortable collecting the information and conducting the content analysis (see Table 2). She collects information from program manuals, written policies and procedures, and the school district website related to the Texas State Plan's Sections 1 and 3. To enhance validity, two other evaluation members independently review the documents, and as a team, they work to come to a consensus before finalizing the report. Given the difficulty of writing good surveys for multiple audiences, outside consultants, in conjunction with the Advanced Academics Office, are hired to design surveys for administrators, teachers, parents, and students. The evaluation team reviews each survey before piloting it and assists in collecting survey data from the various constituent groups. The BDISD Research Office collects and analyzes demographic information comparing the overall student demographics of the district and individual schools' G/T nominations and identified G/T students. The Advanced Academics Office, representatives from the evaluation team, and outside consultants review the identification process and instruments used in identifying students. After collecting lesson plans from G/T elementary teachers and curriculum learning progressions, the Advanced Academics team and the outside consultants review the curriculum. G/T teachers collect TPSP scores and forward them to Advanced Academics to document TPSP baseline scores. Results are shared with the evaluation team throughout the process. The team continually discusses ways to improve interrater reliability and cross-checking to reduce bias of results. Administrators

TABLE 2
Bright Day ISD Data Collection Procedures

Evaluation Questions	Data Collection Steps	Timeline	Who is involved?
1.To what extent are Bright Day ISD's GT programs in compliance with the Texas State Plan (Sections 1 and 3)?	Document review and write draft report of content analysis	October–November	Evaluation Team
	Design and pilot surveys (administrator, teacher, parent, student)	November–January	Outside consultants, BDISD Research Office, and Advanced Academics Office
	Administer surveys	March–April	Evaluation Team and Advanced Academics Office
	Analyze surveys and draft report	May–June	BDISD Research Office and outside consultants
	Review and finalize report	July	Evaluation Team and outside consultants
2a. To what extent do nominated and identified G/T students reflect the district population? 2b. To what extent do nominated and identified G/T students reflect individual campus populations?	Collect demographic data from district and from individual campuses related to nomination and identification and analyze data	October–November	BDISD Research Office
2c. How are instruments used on each campus during the identification process?	Collect instruments and nomination/identification forms used by campuses	October–November	Evaluation Team, outside consultants, and Advanced Academics Office
	Review of instruments, forms used at each campus, and identification process	December	Evaluation Team, outside consultants, and Advanced Academics Office
	Write summary report	January	Outside consultants, BDISD Research Office, and Advanced Academics Office
3a. To what extent do teachers, parents, and students perceive a lack of challenge in G/T classrooms?	Included in survey process above	November–July	Evaluation Team, outside consultants, BDISD Research Office, and Advanced Academics Office
3b. To what extent does the program meet curricular and instructional standards for G/T programs?	Collect lesson plans, curriculum, learning progressions at each campus	October–November	Advanced Academics Office
	Review curriculum		Outside consultants and Advanced Academics Office
	Train principals	February	
3c. To what extent are teachers implementing differentiation in their classrooms? 3d. How are G/T students clustered?	Principals conduct observations	April–May	Advanced Academics Office
3e. What are baseline levels for gifted students' performance on the TPSP?	Collect TPSP projects by campus	March–May	Principals and G/T specialists
	Write summary report	June–July	G/T teachers and Advanced Academics Office Outside consultants and Advanced Academics Office

and G/T specialists receive training in conducting classroom observations and begin these observations after the survey is closed. Observers are not informed of survey results so their perceptions are not influenced by this knowledge.

JUSTIFY CONCLUSIONS

After data collection, the gathered information is analyzed and summarized in a written report. Ideally, the evaluation finds strengths and weaknesses of the program. The preliminary and final reports should include the background information that led to the purpose of the evaluation.

Data analysis and data presentation. The method of analysis depends on whether the collected data were qualitative or quantitative. Qualitative data obtained through interviews, focus groups, or open-ended survey questions will need to be transcribed and coded prior to analysis. Text should be initially reviewed for preliminary themes. Multiple coders/peer reviewers should be incorporated to increase validity. Data can be analyzed with computer software or manually. Quantitative data may need to be cleaned, or checked for accuracy, before analysis (CDC, 2011). Descriptive data should be tabulated for each question and, if appropriate, stratified by demographic variables of interest (e.g., school, age, race). Comparisons between respondent groups are also appropriate.

Interpretation of results. The evaluation findings should be interpreted to determine what the results mean. Potential alternative interpretations should be considered and, if appropriate, provide reasons why alternative explanations should be discounted. The CDC (2011) recommended interpretation based on context (e.g., social context, political context, program goals, stakeholder needs) and discussion of preliminary findings with stakeholders because different parties may have insights that may

guide interpretation, however conclusions must be "drawn directly from the evidence" (CDC, 2011, p. 31). Asking "so what" aids in interpreting findings and suggesting recommendations that are useful to the constituents. Before finalizing the written report, the evaluation team members should familiarize themselves with the evaluation findings and be prepared to discuss conclusions and next steps.

Written report. After data analysis, the evaluation team interprets the data to understand what the findings indicate about the program components. Keep in mind the *purpose* of the evaluation and the intended users of the results. A preliminary report explains the background for the evaluation, the purpose of the evaluation, evaluation questions and method, results, interpretations, recommendations, and limitations of the evaluation. Data should be organized and presented in a logical manner such as by evaluation questions. Conclusions should be justified by (a) limiting conclusions to appropriate contexts, purposes, persons, and time periods; (b) reporting data that supports each conclusion; and (c) including other plausible explanations with reasons for discounting alternatives (Stufflebeam, 1999). Recommendations should be based on the results of the evaluation (CDC, 2011; Milstein et al., 2017). To enhance *usefulness* and readability, an executive summary highlighting the important findings and conclusions should be written last but placed at the beginning of the report.

The evaluation team compiles the evaluation report that includes the recommended action plan (see Table 3). Their final report includes all of the components listed in the program evaluation checklist (see Appendix). This report summarizes parents' concerns of lack of challenge, the principals' concerns of overidentification of students in some schools, and the BDISD School Board's concern with underidentifica-

tion and lack of equitable access to the gifted program. The purpose to collect baseline data is included along with goals and objectives of the evaluation. The report also includes multiple visual components such as graphs, tables, a chart showing how many classrooms lack differentiation, and a flowchart of the current identification process.

The results from the evaluation highlight several inconsistencies in the district gifted and talented program. With regard to compliance, the district shows gaps in the area of student assessment. Although the district uses multiple measures to nominate students, it uses only one measure to identify students for the program. At least 3 assessments should be used in the decision-making process for identification. These assessments need to be collected from multiple sources and include both qualitative and quantitative information. If students are English language learners, the tests may need to be nonverbal or in languages that students understand. Further investigation of student demographic data in BDISD reveals disparities in the number of low-income and ELL students identified for the program across the school district, across campuses, and within campuses at different grade levels. The district does not meet compliance standards in the area of curriculum and instruction because of the lack of challenging learning experiences in foundation curricular areas, low participation rates in development of advanced-level projects, and lack of opportunities to accelerate in students' areas of strength. In the area of service design, the cluster service model adopted by the district varies across campuses. In addition, the stakeholder evaluation team discerns that the gifted students in the clustered classrooms are not provided challenging learning opportunities in more than 60% of the clustered classrooms, and many students are not provided the opportunity to accelerate in areas of strength. The elementary-aged gifted students overwhelmingly perceive the classes as easy and unchallenging; several comments suggest that everyone in the class does the same level of work, and, in some cases, G/T students are used to tutor other students.

TABLE 3
Bright Day ISD's Action Plan

Research Question	Desired Outcome	Gap (Results From Evaluation)	Action to Be Carried Out	Person(s) Responsible	Timeframe
1.To what extent are Bright Day ISD's G/T programs in compliance with the Texas State Plan?	Full compliance with Sections 1 and 3 of the Texas State Plan	Not all campuses are using multiple assessments to identify gifted students	Identify assessments that are being used at nomination and identification phases; examine technical qualities	G/T coordinator and Advanced Academics Office	2 months
		Neither the district nor all campuses' G/T programs reflect demographics of campus	Provide professional development to teachers and administrators in use of assessments	G/T coordinator and outside consultants as needed	During the summer or before the next identification period
		Inconsistencies across campuses exist in terms of curricular and instructional challenge as well as clustering of gifted students	Provide professional development for teachers/mentors in differentiating curriculum and instruction	G/T coordinator and outside consultants as needed	During the summer prior to beginning of school; mentors to support teachers during the year
			Provide professional development for administrators in terms of service design and clustering students	G/T coordinator and outside consultants as needed	Prior to the beginning of next school year
2a. To what extent do identified nominated and identified G/T students reflect district population?	District G/T demographic population should closely mirror the district population	The overall G/T population does not reflect the district population	Identify assessments that are being used at nomination and identification phases; examine technical qualities	G/T coordinator and Advanced Academics Office	2 months
2b. To what extent do nominated and identified G/T students reflect individual campus populations?	All phases of the identification process should reflect the campus populations	Inconsistencies exist across campuses, between nominations and identification, and between grade levels	Professional development will be provided to those involved in the identification process in use of assessments and in providing access	G/T coordinator and outside consultants as needed	During the summer or before the next identification period
2c. How are instruments used on each campus during the identification process?	Each campus uses required-district instruments with fidelity and for their intended purpose	Discrepancies are found in what instruments are used in different identification phases, how they are implemented, and how they are interpreted	Professional development will be provided regarding instruments, how they should be used, and how they should be interpreted	G/T coordinator and outside consultants as needed	During the summer or before the next identification period
3a. To what extent do teachers, parents, and students perceive a lack of challenge in G/T classrooms?	Students in G/T classrooms receive a challenging curriculum commensurate with their abilities and interests	Constituents perceive that G/T classes are not challenging and not addressing student interests	All constituent groups receive an overview of the school district's curriculum and G/T program	G/T coordinator, classroom teachers, principals	During the summer prior to beginning of school

Research Question	Desired Outcome	Gap (Results From Evaluation)	Action to Be Carried Out	Person(s) Responsible	Timeframe
3b. To what extent does the program meet curricular and instructional standards for G/T programs?	Teachers are able to access challenging, differentiated curriculum for G/T students	Fewer than 60% of cluster classrooms have curriculum that is aligned to G/T standards	Teachers receive professional development about designing differentiated curriculum	G/T coordinator, G/T committee	During the summer prior to beginning of school
3c. To what extent are teachers implementing differentiation in their classrooms?	Challenging, differentiated curriculum is taught in all cluster classrooms	Fewer than 60% of G/T teachers in cluster classrooms implement a differentiated curriculum	Teachers receive professional development in how to use preassessments, formative assessments, and a comprehensive curriculum to identify needs and develop differentiated plans	G/T coordinator, G/T committee	During the summer prior to beginning of school; mentors support teachers during the year
3d. How are G/T students clustered?	A minimum of five G/T students are clustered in classrooms with academically advanced students	On some campuses, G/T students are spread across classrooms; on others, clustered with struggling students	All constituent groups receive an overview of the district's service delivery model and research related to grouping and acceleration practices	G/T coordinator, G/T committee	During the spring prior to classroom placements
3e. What are baseline levels for gifted students' performance on the TPSP?	G/T students are performing at the Expert Level on the TPSP at their grade level and show or maintain growth from year to year	The majority of G/T students' baseline performance on the TPSP is at the Apprentice level	Documentation of baseline levels for students to be used as a growth indicator for future years	G/T coordinator, principals, teachers	During the spring prior to classroom placements

SHARE EVIDENCE

Next, the evaluation findings should be communicated to the constituents. The timing and methods for communication, style, tone, format, and the reporting style depend on the audience and their specific needs (Callahan et al., 1995; CDC, 2011). An intentional dissemination strategy should be planned and executed (CDC, 2011). The School Board should receive a short verbal presentation of the findings, the complete evaluation, and the executive summary. Fact sheets or condensed reports may be more appropriate for other constituents, such as students, parents, and teachers. In addition, supporting graphics, stories, and highlighting of important points are helpful.

The evaluation team shares the full evaluation with the Bright Day

Independent School Board and develops an intentional strategy for disseminating the results of the information. Summarized versions of the reports, which include a brief explanation of the evaluation process, are prepared that keep in mind the needs of the various constituent groups (e.g., parents, teachers, students, administrators). Graphs and tables are included to represent the data.

IMPLEMENT ACTION PLAN

The goal of program evaluation is the implementation of an action plan to improve the program. Action plans can delineate the gap between desired outcome and evaluation results, but are most useful when a group of vested parties determines specific action steps, a timeline, and who is responsible for the implementation. The inclusion of stakeholders throughout the evaluation process increases the probability that the results are used (CDC, 2011; Milstein

et al., 2017). Milstein et al.'s (2017) tools section outlined steps for conducting a social marketing campaign to "sell" the action plan to important implementers and create a plan for sustainability. Part of the sustainability plan is to collaborate and attract support that will aid in specific programs. Setting postevaluation meetings with stakeholders after a preidentified point in time (e.g., 6 months) maintains accountability and forward-oriented progress (CDC, 2011).

Colleen and the evaluation team will make recommendations for the desired outcomes pertaining to each research question. If the school board approves the recommendations, Colleen and the Advanced Academics Office will oversee the implementation of the changes to the gifted and

talented program (see Table 3). The evaluation provides baseline data and reveals that the district is not in full compliance with the State Plan in the areas of student assessment and curriculum instruction.

The cluster service model adopted by the district allows gifted students opportunities to work independently but, through observations and results from instruments, the evaluation team discerns that the gifted students in the clustered classrooms are not provided challenging learning opportunities, and many students are not provided the opportunity to accelerate in areas of strength. The team's recommendation to the school board includes providing professional development to help teachers recognize the diverse characteristics of gifted learners and specific training that will assist teachers in differentiated curricular design that allows for a variety of learning opportunities and challenges. Colleen plans dates for training and orientation events for the staff. With the support of administrators and school board, she begins the process of updating policies and procedures for the program. The team recommends examining the technical qualities of the instruments and ways that the campuses are using and interpreting assessments, and suggests that the campuses may want to consider using a nonverbal ability test to assist in identifying nontraditional gifted students (Johnsen, 2011).

Although the process has taken effort, the evaluation advisory team is pleased with its work and can now make strategic decisions that will bring the program into compliance and plan for periodic evaluations in the future.

ENGAGE STAKEHOLDERS

As graphically represented (see Figure 1), evaluation is an ongoing process. In fact, the Texas State Plan (5.3C) requires an annual evaluation of the effectiveness of G/T services. Districts may want to select one of the five strands to address each year for a 5-year rotation, prioritized by district needs. Regardless of which strands are evaluated, the stakeholders need to be continually involved—involved in the ongoing evaluation, involved in the implementation of recommendations, involved in ascertaining if the recommendations from previous cycles are being implemented, etc.

The school board enthusiastically endorses the evaluation team recommendations and schedules meetings in 3 and 6 months to check on the progress of implementation.

CONCLUSION

No matter how large or small a school district, evaluating programs for gifted students is critical to providing high-quality services for students. The evaluation process can be challenging and rewarding for district coordinators and their teams. Evaluating programs involves commitment and follow through of stakeholders. The information collected in the evaluation is invaluable for effective program development. The effort put into the planning will benefit the students now and into the future. Depending upon the results of the evaluation, districts may have simple or extensive action plans. A few concerns may need immediate consideration while others may take more implementation time. In order for change to fully occur throughout the district, administrators and other educators must understand the implications of a strong, defensible gifted and talented program. By developing a robust evaluation and a transformative action plan, the needs of gifted and talented students will match the opportunities available on campuses and dedicated evaluation teams will appreciate the merit of their noteworthy efforts.

REFERENCES

Bohn, C. (2000). *Gifted program evaluation in progress.* Storrs: University of Connecticut, The National Research Center on the Gifted and Talented. Retrieved from http://nrcgt.uconn.edu/newsletters/fall003

Callahan, C. M. (2004). Asking the right questions: The central issue in evaluating programs for the gifted and talented. In C. M. Callahan's (Ed.), *Program evaluation in gifted education* (pp. 1–11). Thousand Oaks, CA: Corwin Press.

Callahan, C., & Hertberg-Davis. (2013). *Fundamentals of gifted education: Considering multiple perspectives.* New York, NY: Taylor & Francis.

Callahan, C. M., Moon, T. R., & Oh, S. (2017). Describing the status of programs for the gifted. *Journal for the Education of the Gifted, 40,* 20–49.

Callahan, C. M., Tomlinson, C, A., Hunsaker, S. L., Bland, L. C., & Moon, T. (1995). *Instruments and evaluation designs used in gifted programs.* Charlottesville: University of Virginia, National Research Center on the Gifted and Talented.

Carter, K. R., & Hamilton, W. (1985). Formative evaluation of gifted programs: A process and model. *Gifted Child Quarterly, 29,* 5–11.

Carter, K. R., & Hamilton, W. (2004). Formative evaluation of gifted programs: A process and model. In C. M. Callahan's (Ed.), *Program evaluation in gifted education* (pp. 13–27). Thousand Oaks, CA: Corwin Press.

Centers for Disease Control and Prevention. (2011). *Developing an effective evaluation plan: Setting the course for effective program evaluation.* Atlanta, GA: Author. Retrieved from https://www.cdc.gov/obesity/downloads/CDC-Evaluation-Workbook-508.pdf

Cotabish, A., & Robinson, A. (2012). The effects of peer coaching on the evaluation knowledge, skills, and concerns of gifted program administrators. *Gifted Child Quarterly, 58,* 160–170.

Fleischer, D., & Christie, C. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation, 30,* 158–175.

Gallagher, J. J. (2006). According to Jim

Gallagher: How to shoot oneself in the foot with program evaluation. *Roeper Review, 28,* 122-124.

Gagnon, D. J., & Mattingly, M. J. (2016). Advanced Placement and rural schools: Access, success, and exploring alternatives. *Journal of Advanced Academics, 27,* 266-284.

Johnsen, S. K. (Ed.). (2011). *Identification of gifted students: A practical guide.* Waco, TX: Prufrock Press.

Johnsen, S. K. (2017). Appendix B. Classroom observation instruments. In S. K. Johnsen & J. Clarenbach (Eds.), *Using the national gifted education standards for Pre-K-Grade 12 professional development* (2nd ed., pp. 231-268). Waco, TX: Prufrock Press.

Long, L. C., Barnett, K., & Rogers, K. B. (2015). Exploring the relationship between principal, policy, and gifted program scope and quality. *Journal for the Education of the Gifted, 38,* 118-140.

Matthews, M. S., & Shaunessy, E. (2010). Putting standards into practice: Evaluating the utility of the NAGC Pre-K-Grade 12 Gifted Program Standards. *Gifted Child Quarterly, 54,* 159-167.

McIntosh, J. S. (2015). *The depth and complexity program evaluation tool: A new method of conducting internal program evaluations of gifted education programs* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

Milstein, B., Wetterhall, S., & CDC Evaluation Working Group. (2017).

Chapter 36. Introduction to evaluation (Sections 1-6). University of Kansas. Retrieved from http://ctb.ku.edu/en/table-of-contents/evaluate/evaluation

Moon, T. R. (2012). Assessing resources, activities, and outcomes of programs for the gifted and talented. In C. M. Callahan & H. L. Hertberg-Davis (Eds.), *Fundamentals of gifted education: Considering multiple perspectives* (pp. 448-457). New York, NY: Routledge.

Olszewski-Kubilius, P., Steenbergen-Hu, S., Thomson, D., & Rosen, R. (2017). Minority achievement gaps in STEM: Findings of a longitudinal study of Project Excite. *Gifted Child Quarterly, 61,* 20-39.

Stufflebeam, D. L. (1999). *Program evaluation metaevaluation checklist (Based on* The Program Evaluation Standards*).* Retrieved from https://wmich.edu/sites/default/files/attachments/u350/2014/program_metaeval_short.pdf

Swanson, J. D. (2016). Drawing upon lessons learned: Effective curriculum and instruction for culturally and linguistically diverse gifted learners. *Gifted Child Quarterly, 60,* 172-191.

Texas Education Agency. (2009). *Gifted talented education: Texas state plan for the education of the gifted and talented students.* Retrieved from http://tea.texas.gov/Academics/Special_Student_Populations/Gifted_and_Talented_Education/Gifted_Talented_Education

Texas Education Agency. (2013). *Gifted talented education: Guiding questions for program review.* Retrieved from https://tea.texas.gov/Academics/Special_Student_Populations/Gifted_and_Talented_Education/Gifted_Talented_Education

Texas Education Agency. (2015). *School dilemmas.* Texas G/T Program Implementation Resource. Retrieved from http://www.texasgtresource.org/school-dilemmas

VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 gifted programs: A clarion call for enhanced gifted program development. *Gifted Child Quarterly, 50,* 199-210.

Walker, L. D., & VanderPloeg, M. K. (2015). Surveying graduates of a self-contained high school gifted program: A tool for evaluation, development, and strategic planning. *Gifted Child Today, 38,* 160-176.

Warne, R. T. (2014). Using above-level testing to track growth in academic achievement in gifted students. *Gifted Child Quarterly, 58,* 3-23.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3).

Corina R. Kaul, M.A., received her B.S. degree from the University of Oregon, her master's degree from Baylor University, and is currently a doctoral candidate in the Department of Educational Psychology at Baylor University where she is specializing in gifted education and quantitative research. She is Assistant Director of the Center for Community and Learning and Enrichment and administers Baylor's University for Young People, an enrichment program for gifted students, and the annual conference for K-12 educators of gifted children. She also assists in conducting program evaluations of district gifted programs. Her current research interests focus on G/T program evaluation, low-income gifted students, gifted first-generation students, teachers of gifted students, and affective needs of gifted learners. She may be reached at corina_kaul@baylor.edu.

Brenda K. Davis, M.A., has more than 20 years of experience as an educator. She received her B.S. from The University of Texas at Austin and her master's degree in professional school counseling from Lindenwood University in Missouri. Currently, she is a doctoral student in the department of educational psychology at Baylor University specializing in gifted education. She serves on the Board of Directors of the Texas Association for the Gifted and Talented. Her interests include gifted education and program design, social and emotional issues of gifted students, and creativity. She may be reached at Brenda_Davis1@baylor.edu

Susan K. Johnsen, Ph.D., is interim chair and a professor in the Department of Educational Psychology at Baylor University where she directs the Ph.D. program and programs related to gifted and talented education. She has written three tests used in identifying gifted students: Test of Mathematical Abilities for Gifted Students (TOMAGS), Test of Nonverbal Intelligence (TONI-4), and Screening Assessment for Gifted Students (SAGES-3). She is editor-in-chief of *Gifted Child Today* and author of more than 250 publications related to gifted education. She is past president of The Association for the Gifted (TAG) and past president of the Texas Association for Gifted and Talented (TAGT). She has received awards for her work in the field of education, including NAGC's President's Award, CEC's Leadership Award, TAG's Leadership Award, TAGT's President's Award, and TAGT's Advocacy Award.

1. Form a Stakeholder Evaluation Team

➤ Consider selecting individuals from each stakeholder group (e.g., principal[s], G/T teacher[s], curriculum coordinator[s], parent of G/T child[ren], etc.)

2. Preplanning

➤ *Purposeful*
- What is the purpose of the evaluation (e.g., formative or summative)?
- Who or what is the impetus for this evaluation?

➤ *Useful*
- *Stakeholders*: Are there any political issues or agendas to keep in mind?
- *Audience*: Who will receive the results of the evaluation?
- *Results*: How will the results be used? Who will use the results?

➤ *Practical*
- What data do we already have that is readily accessible within the district?
- What information do we have about the available budget? Who do we gather this information from?

➤ *Accurate*
- How do we ensure the accuracy of the data?

3. Design and Evaluation

➤ Revisit the Preplanning topics (*purposeful, useful, practical, and accurate*) as the evaluation questions are finalized
- Does the purpose drive the design of the evaluation?
- Will the evaluation be conducted in house or with external assistance?

➤ *Purposeful and Useful*
- What kind of information is needed (e.g., facts, perceptions, processes, beliefs, outcomes, goals)?
- Who will data be collected from (e.g., administration, teachers, parents, students, community)?
- How could data be collected (e.g., document/records review, questionnaire, focus group, interviews, observations)?

➤ *Accurate*
- *Instrument(s):*
  · What instruments will be needed for the review?
  · Is there an existing instrument that could be used?
  · If an instrument is needed, how and who will design the instrument?
  · Is this instrument valid?
- How will data be triangulated? (Get different perspectives on issue to validate perceptions [e.g., ask questions across different groups].)
- How will the data be analyzed?

➤ *Useful*
- Who will use the data collected?
- How will the results be used?

➤ *Practical*
- *Resources*: What resources (e.g., personnel, time, money, IT assistance) are available?
- *Timeline and Personnel*: Create a timeline of items. Who will do what tasks?

4. Gather Evidence

➤ Content analysis—Conduct a document review
- Review relevant Texas State Standards to examine if documents meet standards.
- Other records (e.g., financial records, professional development).

➤ Gather other evidence from other stakeholders (e.g., parents, students, teachers, administrators).

5. Justify Conclusions

➤ Compile and analyze data

➤ Interpret results and write report
- Include background, purpose of evaluation, goals/objectives, evaluation questions, and evaluation methods.
- Be sure evidence supports any conclusions/points made.
- Include visuals (e.g., charts, graphs, tables).
- Indicate positive evaluation findings.
- Include an action plan.
- End on a positive note.
- Consider including limitations, a warning against potential misinterpretations, and potential side effects of the report.
- Write Executive Summary.

➤ Summarize information for various groups

6. Share Evidence

➤ Share preliminary report with evaluation team/key constituents for corrections/verifications of interpretations

➤ Present findings to school board

➤ Develop and implement communications strategy

7. Implement Action Plan

➤ Include desired outcome of actions, gap identified in evaluation process, actions to be carried out, person(s) responsible for actions, and timeframe for implementation

8. Engage Stakeholders

An appendix with Program Evaluation Resources recommended by this article's authors can be downloaded at www.txgifted.org/tempo.