

Pre-AP[®] Science Vertical Teams Summer Institute Evaluation Summary: A One-Year Follow-Up

By Vytas Laitusis and Carol Barry

EVALUATION

Vytas Laitusis is a research scientist at the College Board.

Carol Barry is an assistant research scientist at the College Board.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT[®] and the Advanced Placement Program[®]. The organization also serves the education community through research and advocacy on behalf of students, educators and schools. For further information, visit www.collegeboard.org.

© 2012 The College Board. College Board, Advanced Placement Program, AP, AP Vertical Teams, Pre-AP, SAT, SpringBoard and the acorn logo are registered trademarks of the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners. Visit the College Board on the Web: www.collegeboard.org.

For more information on College Board research and data, visit www.collegeboard.org/research.

EVALUATION

Contents

- Executive Summary 3
- Introduction 4
- Method 4
 - Post-Training Survey..... 4
 - Follow-Up Survey 4
- Results of Post-Training Survey 5
 - Participant Background 5
 - Perceived Institute Effectiveness 5
 - Summary of Open-Ended Responses 6
- Results of Follow-Up Survey 7
 - Participants in Follow-Up Survey 7
 - Institute Effectiveness: Follow-Up Survey 7
 - School/Classroom Environment..... 8
- Comparisons of Institute Effectiveness, Pre- and Post-Follow-Up 10
- Summary of Self-Reported Implementation Changes..... 11
- Conclusion 12

Tables

| | |
|--|----|
| Table 1. Comparison of Post-Institute Survey Responses Between Teachers Included in Follow-Up Sample Versus Teachers Not Included in Follow-Up Sample | 8 |
| Table 2. Question Prompt: How Often Do You Engage in Each of the Following Activities in Your Classroom? | 8 |
| Table 3. Question Prompt: Please Indicate How Frequently the Following Assessment Events Occur in Your Classroom | 9 |
| Table 4. Question Prompt: Please Indicate the Degree to Which Each of the Following Is a Factor in Deciding Student Enrollment in Your School's AP [®] Classes..... | 9 |
| Table 5. Question Prompt: Please Indicate the Extent to Which You Agree with the Following Statements Concerning Your Experiences with the Institute | 10 |
| Table 6. Question Prompt: As a Result of Exposure to Materials and Strategies, Indicate the Extent to Which You Would Change the Following | 11 |

Executive Summary

This study evaluated the effect of a pilot Pre-AP[®] Science Vertical Team Summer Institute on teacher practices and perceptions of the program. The overarching goals of the Vertical Teams Institute were focused on the tools needed to build or strengthen an AP[®] Vertical Team and to align science curriculum across grade levels. The evaluation also surveyed a sample of the initial attendees one year after the institute to assess their perceptions of the institute and practices in the classroom following a year of implementation in the school.

Results:

- Of the respondents, 61% did not have Vertical Teams in their district at the time of the training, but planned to introduce them in the future.
- Of the respondents, 91% indicated that they would change their instructional practices to emphasize certain content, and 95% indicated that they would change their classroom activities.
- Of the respondents, 75% indicated that what they learned would better prepare their students for the rigors of AP.
- Of the respondents, 86% indicated that the strategies and tools they were exposed to would improve student access to the AP program.
- After one year of implementation, there was no significant reduction in the perceived utility of the institute.
- After one year of implementation, there was no significant reduction in perceived impact of the institute on teacher practice.

The primary limitation of this research is that it is based on teachers' self-reports of their intended and actual changes in instructional practice. The follow-up sample, although similar in terms of their initial perceptions of the institute, was small and therefore difficult to generalize.

Introduction

In the summer of 2009 the College Board piloted a series of Vertical Team or Pre-AP Summer Institutes (APSI) in three states throughout the Southwestern region. The overarching goals of the Vertical Teams Institute were focused on the tools needed to build or strengthen AP Vertical Teams® and to align science curriculum across grade levels. The program was also designed to inform participants on how to implement research-based instructional strategies to foster inquiry and reasoning in their students. Participants were expected to leave the institute with inquiry-based model labs and lessons, an aligned set of learning progressions, and a practical action plan for their Vertical Team in order to prepare their students for success in college-level courses.

The institutes were intended for middle and high school science teachers who were involved in forming or strengthening Vertical Teams in their schools. At the conclusion of the institute, participants were expected to have an understanding of how to:

- Build an AP Vertical Team or strengthen an existing one.
- Align science curriculum across grade levels.
- Create and implement an action plan.
- Foster inquiry and reasoning using research-based instructional strategies.

Method

Two sets of surveys were used to collect data on teacher attitudes and behaviors surrounding the professional development (PD). The first survey, *Post-Training Survey*, was a paper-and-pencil instrument administered on-site at the conclusion of the institute. The second survey, *Follow-Up Survey*, was administered online to a subset of the original set of respondents approximately one year after the institute.

Post-Training Survey

Evaluation data was collected from 11 institutes taking place in three Southwestern states from June through August 2009. Each institute lasted approximately 4.5 days. On the final day of the institute, participants were given a survey to complete that was designed to capture their attitudes and perceptions regarding their experiences. All surveys were completed before the participants left the institute, resulting in a total of 212 respondents from the 11 sites.

The survey was broken down into three fundamental sections: participant background, perceptions regarding the quality of the institute's content, and delivery and perceptions regarding the institute's utility. The results presented here focus on the participants' background (for contextual purposes) and the institute's perceived utility or impact with regard to potential changes in classroom practices.

Follow-Up Survey

After participating in the summer institute, teachers had a year to implement the strategies they learned in the institute. The *Follow-Up Survey* was designed to capture their experiences with that implementation. Since the prospective respondents were not in a centralized location for a follow-up administration, a Web-based survey was constructed. The survey

contained a set of items that were matched to the original survey that the teachers had completed immediately after the institute. These common items were used to assess changes in respondent attitudes and practices after a year in their schools. The survey also asked a set of unique questions that tied more directly to respondents' experiences with applying what they had learned to a school setting. In addition, it asked respondents to provide contextual information regarding their teaching practices and use of assessment. All prospective participants were contacted via an email address provided by their original conference registration form. Participants were sent an email asking for their voluntary participation and were assured that their responses, though not anonymous, would be kept strictly confidential. There was no compensation offered for their participation. The direct link to the survey was embedded in the email correspondence.

Results of Post-Training Survey

Participant Background

An overwhelming majority of the participants were teachers, representing 97.6% of the sample. Overall teaching experience ranged from less than one year to 38 years, with a mean of 10.5 years of teaching experience. Only 34 of the participants identified themselves as current AP teachers, with AP teaching experience ranging from one to 13 years, with a mean of 3.4 years. Of the remaining (non-AP) teachers, 25 indicated they would be teaching AP in the coming year. Therefore, 27.8% of the sample represented either practicing or prospective AP teachers.

With respect to their current Vertical Team infrastructure, 23.1% of the participants indicated that they currently have a Vertical Team program established in their school, and 61.3% indicated that they currently do not have a program but plan to in the future. Finally, 11.8% indicated that they do not have, nor do they plan to have, Vertical Teams in their school. Participants were also asked whether they currently use SpringBoard® in their school for Pre-AP, and only 5% of the respondents indicated that they use SpringBoard for this purpose. Approximately half (49.5%) of the respondents indicated that they do not use SpringBoard for Pre-AP, and the remaining respondents said they were not sure.

Perceived Institute Effectiveness

The relative effectiveness of the institute to influence teacher changes in practice was elicited using a set of questions that addressed perceived utility. Teachers were asked to report the level of anticipated change in their practice using a four-point Likert-type rating scale with anchors labeled "not at all," "somewhat," "significantly," and "to a great extent." The most significant anticipated changes reported by respondents concerned teaching practice that was focused on changing content emphasis or priorities, and on changing classroom activities. Of the teachers surveyed in this study, 39.5% indicated that they would change "significantly" or "to a great extent" their content emphasis or priorities (90.7% indicated at least some change), and 49.5% indicated that they would either "significantly" or "to a great extent" change their classroom activities (94.7% indicated at least some change). Although the percentages associated with "significant" or "to a great extent" seem moderate, what they represent may be significant. If teachers had been asked about their *attitudes* regarding the training, one might have expected to see higher percentages of affirmation. However, these questions refer to actual anticipated *changes in behavior*, which is generally much more difficult to effect. When viewed in this light, the percentage of teachers who indicated that they would actually make changes in their schools based on their exposure to this institute

is fairly substantive. Other anticipated changes envisioned in the schools of the participants included (1) changes in curriculum content, which 22.8% indicated they would make “significantly” or “to a great extent”; and (2) grading standards, which 21.1% anticipated they would change “significantly” or “to a great extent.”

In addition to the questions involving *level of anticipated change* in classroom practice, participants were also asked about their *level of agreement* with statements regarding anticipated change in classroom practice. For these questions, participants were asked to rate each of the statements on a four-point Likert-type scale with the following anchors: strongly disagree, disagree, agree, and strongly agree. The set of questions pertaining to classroom behavior utilizing this format focused on whether the knowledge gained from the institute would allow them to better prepare their students for the rigors of an AP classroom, and whether the strategies and tools presented in the institute would increase student access to the AP program. With regard to preparing students for more rigorous course work, 74.9% indicated that they agreed or strongly agreed. With regard to increased access, 86.1% indicated they agreed or strongly agreed that the strategies and tools to which they were exposed would improve access to the AP program.

Summary of Open-Ended Responses

Participants were also asked to respond to two open-ended questions. The first question pertained to additional topics that participants felt should be covered in the institute or in follow-up institutes. There were several themes that emerged from the responses to this question. Several participants indicated that they would have liked more examples of inquiry-based labs, strategies, and classroom activities that they could use in their classrooms (e.g., “I would like to see more inquiry lab and strategies that we can take back to our campus and model to other teachers.”). In contrast, there were also a number of participants who indicated that they would have liked more material on Vertical Teams and less on inquiry-based labs (e.g., “Wish it had been more about vertical teaming and curriculum alignment, and less about inquiry-based instruction.”). Additionally, there were several participants who requested more information regarding the distinction between Pre-AP and AP, as well as between Pre-AP and regular middle school classes. Finally, several participants noted that the Vertical Team Institute would have been more effective had all members of the Vertical Team been in attendance.

The second open-ended question asked for additional information regarding the participants’ experiences with the institute. Again, a number of participants indicated that the institute would be more effective if all members of the Vertical Team attended together (e.g., “It should be stressed to districts participating that this would be much more effective if they had participants from various levels. The purpose of the institute couldn’t be utilized because all levels were not present.”). However, the main theme that emerged in these responses pointed to a misunderstanding or a miscommunication about the content of the institute. For example, some participants expected the institute to cover inquiry-based instruction and labs (e.g., “There was miscommunication as to what type of training I would be receiving. I thought I was getting lab/activities for Pre-AP. I got vertical teaming.”). Others expected the institute to cover vertical teaming (e.g., “I think the title was misleading. We didn’t spend any time at all on vertical teaming” and “This workshop was not about vertical articulation, it was about inquiry. The time actually spent working with our teams was minimal.”). It is possible that this misunderstanding or miscommunication about the institute content explains some of the variability in responses to other survey questions.

A subsequent analysis of the results broken down by institute, not reported here in detail, indicated that there were variations in the quality of institute delivery. Furthermore, these variations in delivery translated to variations in the level of change in the classroom that teachers expected as a result of their participation. These findings highlight the importance of disentangling the professional development (PD) program from the delivery of that program. A PD program's components and the delivery of that program are both distinct yet integrated, and therefore must be accounted for in any analysis of perceived effectiveness.

Results of Follow-Up Survey

Participants in Follow-Up Survey

Contact information for 185 of the original 212 participants was obtained to aid in the recruitment for the one-year *Follow-Up Survey*. Of the 185 email addresses provided by the original conference registration forms, 12 were returned as undeliverable. Of the remaining 173 participants with valid email addresses, a total of 24 responded to either the initial email or a follow-up email that was sent two weeks later. Of the 24 participants who actually logged on to the survey, five had their surveys deleted because they provided little or no survey response data, leaving a total of 19 respondents for subsequent analysis. There was a total of 11 institutes held in three states throughout the Southwest. In the initial sample, a majority of participants took the PD in sites located in one state. However, a majority of the sample who responded to the *Follow-Up Survey* had taken the PD in another state. It is unclear why such a disproportionate number of participants who took the PD in this second state responded to the *Follow-Up Survey*. All of the respondents indicated that they were teachers, and eight of them indicated they either taught AP or AP and another class. The remaining 11 teachers did not teach AP and indicated they taught in either a high school or a middle school.

Institute Effectiveness: Follow-Up Survey

Given the low response rate, an initial concern surrounded the extent to which this follow-up sample was representative of the original survey-taking sample. Table 1 compares the mean responses to an aggregated set of survey items that were collected immediately after the institute (i.e., the *Post-Training Survey*). Each question was based on a four-point Likert-type scale, with "strongly disagree," "disagree," "agree," and "strongly agree" as options. The final set of questions focused on the respondents' views of the PD's estimated impact on their teaching practices, for which response choices ranged from "not at all" to "somewhat," "significantly," and "to a great extent." The questions were aggregated for each section, and the average responses to those questions were compared. The results seemed to indicate a close alignment between the original responses given by the follow-up sample of teachers and the responses of those who did not participate in the follow-up study. These findings suggest that there were no significant biases with respect to the original perceptions of the PD between the follow-up sample and the group that did not participate in the follow-up study. To confirm these observations, all reported means were analyzed using a paired samples t-test, and no significant differences across the two groups were found.

Table 1.

Comparison of Post-Institute Survey Responses Between Teachers Included in Follow-Up Sample Versus Teachers Not Included in Follow-Up Sample

| | Non-Follow-Up Sample | | | Follow-Up Sample | | |
|---|----------------------|------|-----------|------------------|------|-----------|
| | <i>N</i> | Mean | Std. Dev. | <i>N</i> | Mean | Std. Dev. |
| Experience with Pre-APSI Workshop | 189 | 2.87 | 0.66 | 23 | 2.87 | 0.55 |
| Estimated Impact on Practice | 188 | 2.17 | 0.59 | 23 | 2.09 | 0.50 |
| Laying the Foundation | 189 | 3.06 | 0.48 | 23 | 2.99 | 0.60 |
| Thinking Like a Scientist in a Learner-Centered Classroom | 189 | 3.10 | 0.44 | 23 | 3.16 | 0.48 |
| Instructional Strategies for Teaching Science | 188 | 3.17 | 0.46 | 23 | 3.26 | 0.49 |

School/Classroom Environment

In addition to questions surrounding the institute itself, which will be covered in the next section, teachers were asked to characterize their schools or classrooms with respect to pedagogical practices and use of assessment. These responses can be used to build a context for developing an understanding of the interaction of knowledge gained in the institute with their own practice as teachers.

Table 2 summarizes the responses to the question that asked the teachers to rate how often they engaged in certain activities in their classrooms on a five-point scale, with “hardly ever,” “several times a year,” “once or twice a month,” “once or twice a week,” and “almost every class period” as options.

Table 2.

Question Prompt: How Often Do You Engage in Each of the Following Activities in Your Classroom?

| | <i>N</i> | Mean | Std. Dev. |
|---|----------|------|-----------|
| Individualized Instruction | 16 | 4.38 | 0.81 |
| Small-Group Instruction | 17 | 4.29 | 0.77 |
| Teacher-Led, Whole-Group Discussions | 17 | 4.24 | 0.83 |
| Provide Summaries of Key Concepts to Supplement Notes | 16 | 3.88 | 1.09 |
| Group Assignments | 17 | 3.65 | 0.70 |
| Teach Test-Taking Strategies | 16 | 3.31 | 1.20 |
| Lecture | 17 | 3.29 | 1.21 |

Although the responses are self-reported and are not verified by any classroom observations, the activities in which teachers indicate they are the most frequently engaged are consistent with many of the scientific inquiry practices outlined in the institute. However, because there was no classroom practice data collected before participation in the institute, it is also unclear what teachers were already doing before the PD.

Table 3 indicates a similar trend with respect to the use of various forms of assessment in the classroom. The items utilized the same 1–5 rating scale for frequency as the previous question. The responses indicate that these teachers seem to utilize brief and less formal

assessment strategies in the classroom more frequently than other forms of assessment. Verbal probes, which are the cornerstone of formative assessment practices, are cited as the most frequent form of assessment used in the classroom. This frequency is also consistent with the role of formative assessment outlined in the PD institute.

Table 3.

Question Prompt: Please Indicate How Frequently the Following Assessment Events Occur in Your Classroom

| | <i>N</i> | Mean | Std. Dev. |
|---|----------|------|-----------|
| Verbal Probes of Student Understanding | 17 | 4.41 | 0.87 |
| Short Quizzes | 17 | 3.59 | 0.51 |
| Collaborative, Group-Based Products | 17 | 3.00 | 0.94 |
| Structured Self-Assessment | 16 | 3.00 | 1.15 |
| Written Multiple-Choice Tests | 17 | 2.82 | 0.73 |
| Structured Peer Assessment | 17 | 2.76 | 1.25 |
| Longer Semester Tests | 16 | 2.75 | 0.45 |
| Independent Project-Based Work Products | 17 | 2.35 | 0.70 |
| Oral Quizzes | 17 | 2.65 | 1.54 |
| Written Test Requiring Student-Produced Responses | 17 | 2.65 | 0.79 |

Finally, with regard to school/classroom environment, the teachers were asked to rate the importance of a list of mechanisms that can be used to determine a student's inclusion in or eligibility for AP. Table 4 summarizes each of these potential mechanisms with ratings from 1–3, whereby 1 = not a factor, 2 = minor factor, and 3 = major factor. Although none of the response choices received an overwhelmingly high rating, recommendations by school counselors, administrators, and parents seemed to be the most salient mechanisms for identifying students. The PSAT/NMSQT® score received the lowest perceived weighting by the teachers surveyed. Given the low rating of the PSAT/NMSQT, it is unclear whether the respondents were characterizing the PSAT/NMSQT as a barrier or screening tool as opposed to a tool for identifying typically underrepresented students who may not enter AP courses through the usual recommendation channels.

Table 4.

Question Prompt: Please Indicate the Degree to Which Each of the Following Is a Factor in Deciding Student Enrollment in Your School's AP Classes

| | <i>N</i> | Mean | Std. Dev. |
|---|----------|------|-----------|
| School Counselor/Administrator Recommendation | 11 | 2.45 | 0.69 |
| Parent/Guardian Recommendation | 11 | 2.36 | 0.81 |
| Self-Nomination | 10 | 2.30 | 0.82 |
| Completion of Prerequisite Course | 11 | 2.27 | 0.79 |
| Teacher Recommendation | 11 | 2.18 | 0.87 |
| Grade in Prior Course | 12 | 2.17 | 0.83 |
| PSAT/NMSQT Score | 11 | 1.18 | 0.40 |

Comparisons of Institute Effectiveness, Pre- and Post-Follow-Up

The final two tables summarize the differences in teacher responses to a common set of survey items that were included in the original *Post-Training Survey* and then again in the *Follow-Up Survey*. Therefore, these comparisons are based on the same participants over two time periods (as opposed to the analyses reported in Table 1, which was a comparison across two groups for a single time period). The comparisons of original responses to those taken after a one-year follow-up are designed to elicit some understanding of how perceptions surrounding a PD experience changes after an attempt is made to implement what is learned in the school or classroom.

Table 5 summarizes the self-reported experiences of the teachers immediately after the institute and then again after the one-year follow-up. Teachers were asked to indicate their level of agreement on a series of statements using a four-point Likert-type scale, with “strongly disagree,” “disagree,” “agree,” and “strongly agree” as options. With the exception of recommending the institute to a new teacher, mean ratings for all other statements went down slightly. It should be noted that these differences were not statistically significant, which could be due to the very small sample size ($N = 16$). Therefore, it is unclear whether this pattern represents a true decline in rating or an aberration attributed to the size and composition of the sample.

| Table 5. | | | | | | |
|---|---------------|-----------|-----------|-----------|-------|-----|
| Question Prompt: Please Indicate the Extent to Which You Agree with the Following Statements Concerning Your Experiences with the Institute | | | | | | |
| (n=16) | Post-Training | | Follow-Up | | t(15) | p |
| | Mean | Std. Dev. | Mean | Std. Dev. | | |
| Recommend to New Teacher | 2.88 | 0.72 | 3.06 | 0.93 | .64 | .53 |
| Recommend to Experienced Teacher | 2.94 | 0.44 | 2.75 | 1.00 | -.89 | .38 |
| Material Relevant | 2.81 | 0.66 | 2.63 | 0.81 | -.82 | .42 |
| Allow for Better Preparation of Students | 3.00 | 0.37 | 2.56 | 0.81 | -1.96 | .07 |
| I Will Use Strategies to Increase Access | 3.13 | 0.50 | 2.75 | 0.86 | -1.69 | .11 |

Table 6 summarizes the set of questions that asked the teachers to rate the level of impact the PD will have (*Post-Training Survey*) or has had (*Follow-Up Survey*) on their practice. In the *Post-Training Survey*, teachers were asked to estimate how the PD might impact their practice; in the *Follow-Up Survey*, they were asked the same question in reference to how the PD actually *did* impact their practice. Teachers used the same four-point Likert scale pertaining to the same level of agreement used in Table 5. Once again, with the exception of one activity (grading standards), teachers rated the activities lower compared to the initial *Post-Training Survey*. However, once again, the differences were not statistically significant.

Table 6.

Question Prompt: As a Result of Exposure to Materials and Strategies, Indicate the Extent to Which You Would Change the Following:

| | N | Post-Training | | Follow-Up | | t | df | p |
|--------------------------------|----|---------------|-----------|-----------|-----------|-------|----|-----|
| | | Mean | Std. Dev. | Mean | Std. Dev. | | | |
| Curriculum Content | 15 | 2.07 | 0.46 | 1.73 | 0.46 | -1.78 | 14 | .09 |
| Content Sequence | 15 | 1.80 | 0.56 | 1.53 | 0.64 | -1.29 | 14 | .21 |
| Content Emphases or Priorities | 15 | 2.27 | 0.46 | 2.13 | 0.64 | -.69 | 14 | .49 |
| Classroom Activities | 16 | 2.38 | 0.50 | 2.00 | 0.89 | -1.56 | 15 | .13 |
| Homework Expectations | 15 | 1.87 | 0.64 | 1.73 | 0.70 | -.56 | 14 | .58 |
| Grading Standards | 16 | 1.94 | 0.68 | 1.94 | 0.77 | 0 | 15 | 1.0 |

The lack of significant differences reported in Tables 5 and 6 indicates a consistent and positive set of ratings from teachers regarding their institute experiences and the institute’s perceived effect on their practice. However, although the differences in the results from the *Post-Training Survey* and the *Follow-Up Survey* were not significant, again, perhaps because of the small sample size, the ratings on the *Follow-Up Survey* did tend to be slightly lower. There are plausible reasons why ratings could be lower on the *Follow-Up Survey*. The pattern may point to an actual consistent decline in response ratings after one year of implementation, with the lack of significance reported in Table 5 (statements regarding overall experiences with institute) and Table 6 (statements regarding changes to practice) attributed to the size of the sample or the power of the analysis. An explanation for the decline could center on the inherent difficulty in accounting for all the challenges or obstacles that typically confront the implementation of a set of new strategies. It is not until the use of these strategies is actually attempted in a school setting that the teacher actually sees firsthand the challenges associated with this endeavor.

Summary of Self-Reported Implementation Challenges

The teachers were asked during the *Follow-Up Survey* to indicate whether they had engaged in any Vertical Team activities during the year. Of those responding, only 38% indicated that they had engaged in Vertical Team activities. If the teachers indicated that they had not engaged in Vertical Team activities, they were asked to indicate why. The question also had a follow-up that asked them to indicate the reasons for the lack of engagement. There was no consensus regarding the lack of Vertical Team activity, although some cited funding, Vertical Teams still being under development (i.e., alignment work still ongoing), lack of interest by fellow teachers and administrators, and/or a belief that Vertical Teams are only for high school teachers as causes.

As a follow-up, teachers were also asked to identify barriers to implementing Vertical Teams. Responses to this question yielded three major themes. The first barrier centered on school compatibility. Some teachers indicated that although they bought into Vertical Teams and scientific inquiry as concepts, some of their colleagues and administrators with whom they would need to collaborate did not share this same level of buy-in. It is unclear whether this disconnect was due to a difference in philosophy or a difference in training. The second, and related, issue involved training inclusion. Some respondents felt that their whole Vertical Teams unit would have benefited from the training, and that this experience would place them all on the same page with regard to implementation. This inclusion of team members extended to administrators (i.e., principals) because they often set the tone and create affordances (release time, funding, etc.) in the system to allow these activities to take place.

This leads to the third interrelated barrier for Vertical Team success: time. Even with proper training, teachers must be given the opportunity to engage in these activities. Administrators demonstrate their support for change by allowing teachers to engage in PD, but it is equally important to follow through by creating an environment in which the newly learned strategies can be implemented properly.

Conclusion

The results summarized here seem to indicate an initial positive effect with respect to anticipated changes in classroom practice as a result of participation in the institute. Although the findings indicate that many teachers plan to change their practices as a result of this PD experience, it is also important to verify whether these changes actually take place. Two methods of inquiry that can be adopted to address this concern are classroom observations and follow-up teacher surveys. *The Follow-Up Survey* focused on the latter.

Results of the *Follow-Up Survey* seem to indicate that the positive ratings that the institutes received immediately after the PD were maintained after one year of implementation in the school. However, caution regarding the stability and generalizability of these results is warranted because most of the ratings were lower and because the sample size for the follow-up was relatively small.

Teachers seemed to vary somewhat with respect to the successful implementation of Vertical Teams in their schools. It is important for any PD to recognize any potential barriers that teachers might face so that they can be better prepared to face these challenges when they go back to their school. This can be accomplished by addressing these challenges at the initial point of PD contact and, more importantly, through ongoing PD that engages participants multiple times throughout the year(s).

Another challenge some participants of this particular APSI faced was confusion over the focus of the PD. Some felt there was too much emphasis on Vertical Teams at the expense of scientific inquiry, and others felt the opposite. If a similar PD model were to continue, it may be more effective to divide this institute into two distinct courses. Another issue raised by some respondents was the desire to have the entire Vertical Team group attend the PD. Some felt they did not obtain the necessary support from colleagues when they returned. Having all members of the team would greatly increase the cohesiveness of the team from that standpoint.

Although the *Follow-Up Survey* yielded some potentially intriguing results, there were some limitations associated with its methodology. The first limitation concerns the lack of information regarding teacher practice before the institute. Having these data would allow for a clearer investigation of impact when viewing the change in these metrics over time and experience. Another shortcoming concerns the sample size. It is often difficult to make significant conclusions from such a small sample, especially when the sample is not uniformly distributed across the original group. When such a small percentage of an original target responds, there are significant risks regarding the representativeness of this group. Although they seemed similar across some of their initial perceptions of the PD, there may have been other relevant differences between the two groups that were not captured. Additionally, although the survey is a valuable tool for collecting information on what is happening in a school, other forms of data such as observations, document artifacts, or third-party verification (from, for example, an educational manager or district liaison) would provide a clearer and more complete picture. Additionally, this evaluation was based on a set of pilot institutes. However, even if the institute were not to continue in its current format, these results should still inform components of the institute that will be carried forward and for PD in general.

The Research & Development department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP[®], SAT[®], PSAT/NMSQT[®]
- Developing standards and conducting college and career readiness alignment studies
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

| | |
|------------|-------------|
| Admission | Measurement |
| Alignment | Research |
| Evaluation | Trends |
| Fairness | Validity |

