COLLEGE BOARD ADVANCED PLACEMENT® BEST PRACTICES COURSE STUDY

Prepared for
College Board

June 30
2006

Principal Investigator
David T. Conley, Ph.D.

Primary Project Staff
Kirsten Aspengren
Katharine V. Gallagher
Odile Stout
Darya Veach
Damian Stutz

Center for Educational Policy Research (CEPR)

Dr. David T. Conley, Director
University of Oregon
720 East 13th Street, Suite 202
Eugene, OR 97401
Phone 541.346.6153 • Toll-free 877.766.2279
http://cepr.uoregon.edu

Table of Contents

Figures and Tables

Center for Educational Policy Research (CEPR)

The Center for Educational Policy Research seeks to help policy makers and policy implementers alike do a better job of using educational policy as a tool to improve schooling and student learning.

CEPR works with federal agencies, state education departments, non-governmental organizations, private foundations, and school districts to support research on a range of issues in the areas of high school-to-college articulation, adequacy funding, large-scale assessment models, and other policy initiatives designed to improve student success.

On this study, CEPR worked in partnership with the Educational Policy Improvement Center (EPIC), a 501(c)3 not-for-profit organization with expertise in validity studies and high school-college articulation issues.

To cite this report:

Conley, D., Aspengren, K., Gallagher, K., Stout, O., Veach, D. (2006). College Board Advanced Placement® Best Practices Course Study. Center for Educational Policy Research, Eugene, Oregon.

1 Overview

1.1 College Board Advanced Placement® Best Practices Study and CEPR

The Center for Educational Policy Research (CEPR) at the University of Oregon, in partnership with the Educational Policy Improvement Center (EPIC), conducted the College Board Advanced Placement Best Practices Course Study, a project sponsored by the College Board. The study examined the content and structure of college courses that demonstrated "best practices" in seven subject areas tested in the College Board's Advanced Placement (AP) program: biology, chemistry, physics, environmental science, European history, US history and world history. Researchers first developed criteria for the best practices and then identified courses that met those criteria. Results from this study will inform the College Board commissions' work when they review and make recommendations for the redesign of AP courses in these subject areas. The goals of the commissions are to ensure that high school AP teachers emphasize the proper content focus and, more importantly, systematically help students develop the crucial attitudes and skills necessary to thrive in a college classroom.

1.2 Background

The present program of high school AP courses and examinations has been viewed as largely successful by educators, parents, and others, as evidenced by the steadily increasing number of students enrolled in AP courses and taking AP exams. The basis of the AP program is to provide an experience to high school students equivalent to what they will encounter in a typical introductory course taught at colleges and universities nationwide. Given the increasing importance of the AP curriculum in high schools and its popularity with students, it is incumbent upon AP to ensure that all courses taught under the AP rubric reflect the best of college courses. This helps ensure that students who take AP courses are properly prepared for college success and that high school teachers gear their teaching of AP to the best college practices.

The College Board periodically reviews all AP courses and makes modifications as necessary to reflect changes that occur in the subject area and how the subject is taught in college. Previously, those reviews gained their data from curriculum topic surveys distributed to college instructors at a wide range of institutions. The design of this study goes well beyond a topical survey in order to investigate what is taught and what is important in courses identified as being examples of best practices at the institutions that nominated them. This study is conducted in support of the College Board's intent to ensure that AP courses focus on what is most important for college success. In other words, the College Board wishes to ensure that AP courses result in deeper and richer learning and understanding by using as a reference point for redesign college courses that reflect best practices.

1.3 Study Design Overview

The main goal of the study is to provide information to the College Board's AP redesign commissions that supports their efforts in the redesign of AP courses and related professional development for teachers. In order to accomplish this goal, this study provides the following information:

- A set of empirically derived criteria that clearly delineate what is most important for students to know and be able to do in a best practices college course.
- A set of best practices courses that align with these criteria in each subject area.
- In selected subject areas, a composite course that narrates the critical attributes from the set of best practices courses and exemplifies the concepts, principles, and techniques as well as assignments and tests of courses identified as possessing best practices attributes.

In order to generate these findings, the study drew upon data that described and identified best practices in each subject area. First, the study designed a criterion-based instrument to rate best practices courses. Then, the study recruited institutions that enroll the most students who took AP examinations in each subject area and asked them to nominate courses that were examples of best practices. For each nominated course, two primary sources were analyzed: 1) ratings by instructors against the criterion-based rating instrument and 2) ratings by external raters who used the same criterion-based instrument to analyze course documents, such as syllabi, assignments, and tests, as submitted by instructors.

Researchers developed a multi-step process to identify best practices courses and their attributes. The process contained the following major components:

1. Engage content experts to develop seven subject-specific, criterion-based instruments that identified the content knowledge, habits of mind, and instructional practices that should be found in best practices college courses in each subject area.
2. Identify a pool of courses nominated as best practices by higher education institutions that receive the most applications from high school students who have taken AP examinations in each subject area being studied.

3. Have instructors of nominated courses and external raters who are also subject area experts complete the instrument in order to identify from the pool of potential best practices courses in each subject area those courses that best exemplify best practices.
4. Utilize expert panels to review the candidate best practices courses and validate whether these courses do in fact represent best practices in each subject area.
5. Analyze these courses to determine the commonalities that exist among them and the critical attributes that make them best practices courses.
6. Transmit this information, along with the instruments developed to identify best practices courses, to commissions being convened by the College Board in the summer of 2006 and charged with the redesign process for each subject area.

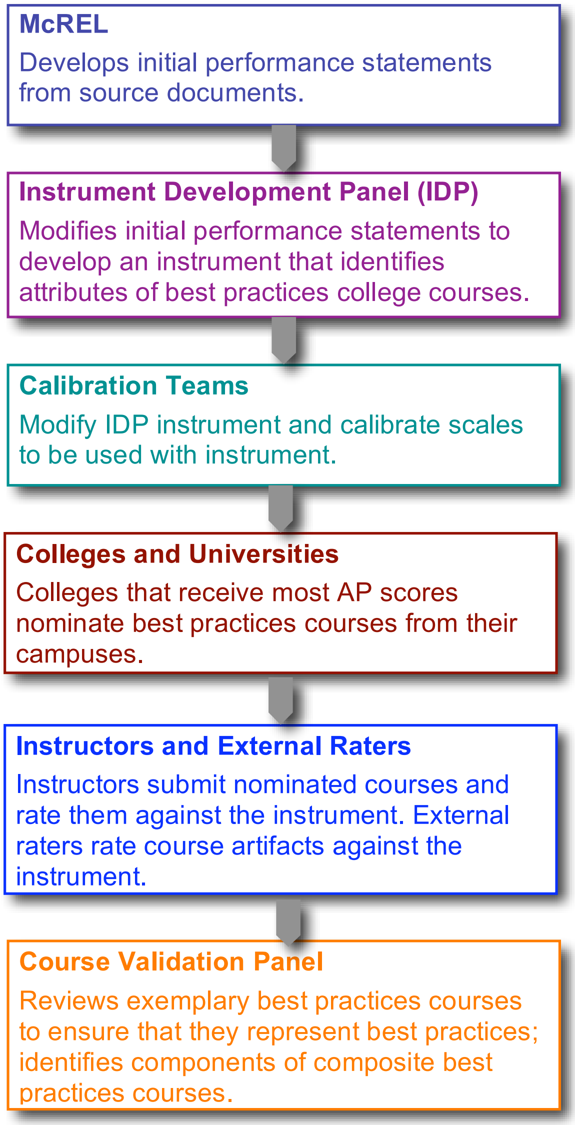
2 Research Design and Methodology

2.1 Overview

From a methodological perspective, this study can be broadly characterized as a validity study. It seeks primarily to establish what constitutes appropriate content, intellectual skills, and instructional practices for AP courses and exams so that valid inferences about students' abilities to successfully undertake college-level work can be drawn from their AP exam scores. The most generally accepted method for identifying appropriate content is the use of expert judgment models. These judgments are commonly used to help inform the construction of graduate admissions exams in a range of professional fields and for national certification and licensing exams. They are most appropriate when specific content knowledge can be identified as being important to subsequent success in an area of study or certification.

The primary methodological technique for the study is best described as a modified version of the Delphi method, also known as convergent consensus. The basic principle of this methodology is to recruit experts who independently identify an initial set of outcomes—in this case, elements of best practices college courses—then successively review and refine those outcomes in order to describe such courses in ever-increasing detail through a process of successive judgments. The method, as adapted for this project, employs a multi-step process to achieve what might be considered sequential or nested convergent consensus. This involves achieving consensus within steps in the process and then across steps. The result is that outcomes (the criteria of best practices courses in seven subject areas) are continually refined and confirmed by multiple independent groups of experts. Figure 1 provides a graphical view of the multiple measures employed in this study.

Figure 1: Research design measures



McREL

Develops initial performance statements from source documents.

Instrument Development Panel (IDP)

Modifies initial performance statements to develop an instrument that identifies attributes of best practices college courses.

Calibration Teams

Modify IDP instrument and calibrate scales to be used with instrument.

Colleges and Universities

Colleges that receive most AP scores nominate best practices courses from their campuses.

Instructors and External Raters

Instructors submit nominated courses and rate them against the instrument. External raters rate course artifacts against the instrument.

Course Validation Panel

Reviews exemplary best practices courses to ensure that they represent best practices; identifies components of composite best practices courses.

2.2 Instrument Development–Best Practices Criteria

In order to identify best practices college courses, it was necessary to develop instruments capable of distinguishing such courses from all other courses in each of seven distinct subject areas. To accomplish this, the following multi-step process was followed in each subject area.

- Development of baseline performance statements
- Development of the instrument
- Calibration of the instrument

2.2.1 Development of Baseline Performance Expectations

Mid-continent Research for Education and Learning (McREL), a nationally known educational research laboratory with expertise in content standards, developed baseline Performance Expectations[1] (PEs). These initial statements were intended solely to provide a starting point for the convergent consensus process. To develop these initial statements, McREL consulted numerous sources, including national reports in the respective subject areas and McREL's own extensive database of standards. (Appendix A - lists documents consulted by McREL during instrument development process.) When synthesizing the content, McREL employed explicit criteria in order to capture the breadth and depth of the content across all sources. If the content was mentioned in two or more sources, it was reworded into a statement of consistent phrase length and language style. The criteria McREL employed when identifying content included the following: 1) appeared frequently in national documents from subject area groups recommending reforms or improvements in the subject area; 2) appeared frequently in state content standards from states that were identified as having the highest quality and most rigorous content standards; and 3) were in the McREL database as important content knowledge, based on previous analyses of content standards in the subject area.

[1] The term Performance Expectation was changed from the original designation of Performance Statement after the Calibration teams review.

For each subject area, McREL developed initial PEs in three sections: Topical Frameworks, Habits of Mind, and Instructional Practices.[2]

- **Topical Frameworks** are the content material covered in each course. The frameworks serve to identify the content that is necessary for students to master in high school AP courses. Some content is more important than other content. The goal of constructing topical frameworks is to allow AP courses to be redesigned to focus on the most important content while not ignoring necessary supporting content.
- **Habits of Mind** are the ways of thinking that students are expected to develop throughout a course of study. Examples include critical thinking, analytical thinking, and inquisitiveness. These habits of mind are of equal importance to the content knowledge specified in the topical frameworks.
- **Instructional Practices** are the specific and general techniques and policies employed by instructors teaching introductory-level college courses. This category includes, for example, teaching methods, assessment practices, policies on student involvement, and uses of technology to assist student learning.

Content in each of the three sections was broken down further into subsections, referred to as Standards, which in turn had multiple Strands identifying major topics within a Standard. PEs were then grouped by strand within a standard.

2.2.2 Instrument Development Panel

Seven Instrument Development Panels (IDPs), one in each subject area, were constituted to 1) conduct the initial review of the baseline PEs and 2) develop detailed criteria to clearly delineate what is most important for students to know and be able to do in best practices college courses. The goal was to produce an instrument that college faculty and trained external raters who were also college faculty could use to report on the content of courses that were nominated as best practices examples. These criteria were established through five successive reviews by the IDPs, with each review designed to achieve greater consensus as well as to incorporate new ideas and allow for edits and alterations. The final result was seven distinct instruments designed to identify best practices courses in each subject area.

[2] The term Instructional Practices was changed from the original designation of Teaching Methods after the Calibration teams review.

2.2.2.1 IDP Recruitment

Following is a detailed description of the methodology employed to identify the content area experts who participated in the IDP online review process.

Step 1: Identify Distinguished Faculty to Nominate Subject Area Experts

Researchers began by conducting an extensive web search of over 850 postsecondary institutions to collect names of distinguished faculty in each of the seven subject areas. These institutions were drawn from those invited to participate in the study. The search resulted in 453 distinguished faculty who were asked to nominate outstanding faculty to be members of the IDP for their subject area.

Faculty were considered distinguished if they met at least one of the following criteria: 1) prestige/status or acknowledged expertise in their subject area, 2) author of publication/report related to improving undergraduate teaching, 3) member of subject-related national group/committee/organization, or 4) recipient of subject-related awards.

The web research yielded a total of 453 distinguished faculty across the seven subject areas. The breakdown for each subject area was: biology 90, chemistry 111, physics 65, environmental science 32, European history 43, US history 62, and world history 50.

Step 2: Contact Distinguished Faculty

Researchers sent a letter via email to the distinguished faculty providing background on this study and asking faculty to nominate leading experts in their fields or to nominate themselves. In addition, they were asked to identify organizations involved in improving undergraduate courses in the subject area that might have individuals who are interested in participating in the IDP online review process and, potentially, on a Course Validation Panel (CVP).

Distinguished faculty were asked to nominate people who met one or more of the following criteria:

1. Individual was acknowledged expert in the subject area.
2. Individual demonstrated leadership in improving undergraduate courses in the subject area.
3. Individual was a member of national organizations or task forces that emphasize or study improvements in undergraduate courses in the subject area.
4. Individual had received awards or other recognition for improving courses in the subject area.

Step 3: Evaluate Nominations

Researchers employed the following nomination tracking procedures to facilitate the evaluation of each nominee's suitability for participation:

- Researched the name, institution, and institution's Carnegie classification for each nominated expert
- Tallied the frequency with which a person was nominated by distinguished faculty
- Reviewed each nominee's curriculum vitae

Researchers contacted all nominees by email to notify them of their nomination and to ascertain their interest in being considered for participation on an IDP. Subsequent phone calls with interested nominees addressed the following points: overview of the study, role of subject area experts in the IDP online review process, confirmation of nominee's qualifications to serve on an IDP, and determination of nominee's interest and availability.

Following each phone call, researchers drafted a short summary of the nominee's qualifications and made a final recommendation about whether to extend an invitation. (Appendix B - lists Instrument Development Panel panelists' qualifications

Step 4: Extend Invitations to Subject Area Experts

Using frequency counts to rank nominees, along with information garnered through phone interviews, researchers ranked subject area experts and extended invitations to the top candidates in each subject area to participate in the instrument development process. The invitation explained what participation entailed, recognition an expert could expect to receive, and the compensation being offered. Researchers followed up on the invitation with telephone contact to answer any questions.

When necessary, researchers used the ranking list to extend additional invitations to nominated subject area experts until participation by at least ten qualified subject area experts was secured for each of the seven IDPs. (See Appendix B for a list of the panelists.)

2.2.2.2 Online Review Process

The IDPs participated in a five-stage, subject-specific online review process designed to reach consensus on the breadth and depth of course content that should be present, the habits of mind students should be developing, and the instructional practices that should be employed in a corresponding, best practices course.

To maximize participation and better accommodate panelists' schedules, the five reviews were conducted using a custom-designed online web tool. The tool was accessible for a limited amount of time, typically seven to ten days. The reviews were designed to allow panelists to provide several different kinds of feedback. Each panelist was given the opportunity to: 1) suggest changes to PEs through multiple iterations, 2) select the top choice in wording of the PE written in its entirety, and 3) assign a verb to a PE indicating how students would best demonstrate knowledge of that content.

Of the 83 panelists across the seven subject areas, a minimum of 93% completed each review. Following are the details of the five reviews. (See http://cepr.uoregon.edu/cbap.start.php for a more detailed description of the online review process for each subject area.)

Review 1

Panelists examined the baseline PEs from McREL, suggested edits, and rated the importance level of each. When suggesting changes to PEs, panelists were instructed to employ one of four types of edits.

1. Simplify the statement, breaking it down into component concepts
2. Expand the statement, giving it more detail
3. Comment on the content of the statement
4. Combine related statements

Panelists were also given the option to suggest new PEs to be added in each of the three sections. Although all suggested changes were taken into consideration, edits and new PE suggestions were not guaranteed to be included in future reviews.

McREL synthesized the 3,000 suggested edits and new PEs, utilizing a set of decision rules to determine which revisions to incorporate into review 2. McREL provided a rationale for each decision. Revisions were incorporated when they met all of the following conditions:

- Had a strong communication value (made the PE easier to understand)
- Had specificity
- Addressed an important issue that was overlooked
- Had support from three or more experts

In addition to suggesting changes, panelists were instructed to rate each PE with regard to its importance as a criterion for a best practices course. Importance level ratings were made using the following four-point scale: 4-Most Important; 3-Important; 2-Less Important; and 1-Least Important.

To help distinguish content that was most important, importance ratings were averaged and minimally reconstructed to give each PE a single importance value. To do this, a weighted average was calculated for each statement by multiplying each rating value by the number of panelists who gave each rating. The sum of this product was then divided by the total number of panelists in each subject area. Use of the weighted average method allowed consensus among the panelists to be more accurately captured.

For reporting purposes throughout all reviews, the four-point importance scale was converted to a three-point scale by collapsing the lowest two levels into one, as follows: 4-Most Important, 3-Important, 2-Less/Least Important. This allowed the statements that were potentially most critical to best practice to rise to the surface and be clearly and quickly distinguishable from those that may not have been quite as crucial. Using the collapsed three-point scale, the weighted average was applied to the following decision rule to determine a single importance rating for each performance statement:

4.0 - 3.5: Most Important
< 3.5 - 2.5: Important
< 2.5: Less/Least Important

Although they were combined for reporting purposes, "Less" and "Least" remained as separate points on the rating scale in future reviews to allow for more fine-grained differentiation of priorities.

Review 2

Panelists reviewed all statements to ensure that they comprehensively and collectively described a best practices course. Panelists were instructed to further edit and add new PEs, as well as to identify importance ratings for edited and new PEs as developed in review 1. For each section, PEs that required no rewriting were grouped and presented by importance level, maintaining their standard and strand structure within each importance group. Panelists were given the opportunity to comment on the PEs collectively within each group.

New and revised PEs were also presented and were organized by their original standard and strand structure. As much as possible, revised and new PEs were displayed alongside their baseline counterparts for comparison. The importance level determined from review 1 was presented at the end of each edited PE. Panelists were given the final opportunity to make further edits to a PE or add new PEs. They were also required to rate each new and revised PE on the four-point scale employed in review 1, providing arguments when their importance rating differed from that presented from review 1.

Data from review 2 were analyzed in a similar fashion to those collected in review 1. PEs rated on the four-point scale of importance, and not requiring further rewriting, were scored and categorized into the three importance level groups. Panelists' arguments for changing a PE's importance level were presented in review 3.

McREL synthesized suggested edits and new PEs. Decisions were made with regard to the inclusion/omission of the suggested revisions according to the set of decision rules employed in review 1. Once again, McREL provided a rationale for each decision. Any written comments pertaining to larger issues were presented to the calibration team (see the next subsection, below, for a description of the team's responsibilities).

Review 3

Review 3 required panelists to achieve final consensus on PEs that comprehensively and collectively described a best practices course. No edits or new PEs were permitted, though substantive comments about what should be learned in a best practices course were encouraged. For each section, PEs that required no rewriting or importance level changes in review 2 were re-grouped and presented by importance level, maintaining their standard and strand groupings within each importance level.

Revised PEs or those with arguments for changing importance levels were organized by their review 2 section, standard, and strand groupings. The importance level from review 2 was presented at the end of each PE followed by any arguments for changing importance levels. Panelists were required to review their colleagues' arguments and, again, rate each PE on the four-point scale.

Data from review 3 were analyzed in a similar fashion to those collected in previous reviews—PEs were scored and categorized into the three importance level groups. Final groupings of PEs for reviews 4 and 5 were achieved using these importance level weighted averages.

Again, any written comments pertaining to larger issues were presented to the calibration team. As edits and additions were no longer accepted, McREL was not required to analyze written input.

Review 3 resulted in complete and final collections of habits of mind and instructional practices; topical frameworks statements were revisited in reviews 4 and 5.

Review 4

Panelists further developed the learning required in a best practices course by identifying preferred language for PEs in the topical frameworks section. PEs were organized and presented by the importance level determined in review 3, maintaining their standard and strand groupings within each importance level. Panelists were required to rank order their top three choices of action verbs with which to associate each statement. Verbs were divided into four cognitive levels, developed by the principal investigator. The four levels are:

- **Level 1**. Retrieval: Know/Understand that, Describe, Provide a historical perspective on, Measure or Calculate, Use
- **Level 2**. Comprehension: Understand/Know how, Explain
- **Level 3**. Analysis: Predict
- **Level 4**. Utilization: Develop a solution using, Design

Researchers advanced the top three verbs in each subject area that received majority approval to review 5.

Review 5

Panelists were required to achieve final consensus upon the action verbs best associated with topical framework PEs. The top three verbs for each PE as determined in review 4 were rank-ordered by panelists. In order to continue moving toward consensus, panelists were not permitted to suggest new verbs. The final action verbs selected were those achieving a minimum 80% level of agreement within each panel. Researchers then created each subject's final instrument by grouping PEs in sections, standard and strands.

2.2.3 Calibration Teams

Once the seven subject-specific instruments were developed, the next task was to determine how the various elements in each instrument would be weighted. This involved assigning relative values or weights to the multiple levels of the instruments (sections, strands, and PEs) as well as between the two data sources (instructor ratings and external rater ratings). To accomplish this, calibration teams varying in size from four to six people were constituted in each subject area. Each team was charged with fine-tuning its respective instrument to ensure that the instrument detected differences between regular and best practices courses. This process also involved practice scoring of actual course material to validate instrument sensitivity.

2.2.3.1 Calibration Team Recruitment

AP staff worked with disciplinary associations and stakeholders to identify subject area experts who could assign values to the instrument so that each instrument would accurately detect a best practices course. Researchers received lists of names from AP staff, rank ordered for each subject area. Researchers recruited calibration team members from these lists and, when necessary, additional sources. Some of these individuals had also served previously on an IDP or served subsequently on a CVP, which helped to create continuity across tasks because these individuals could answer questions regarding other stages in the process in which they had participated. Teams were comprised of five members, with the exception of US history (six) and chemistry (four). (See Appendix C - lists Calibration Team members institutions, associations or high schools.)

2.2.3.2 Pre Calibration Meeting Work

To familiarize calibration team members with the instrument and to generate scoring results for use in calibrating the instrument, each member was required to log on to a specially developed online instrument to review and rate three “mock” courses for evidence of the PEs before attending the calibration team meeting. The mock courses were developed by subject area experts and consisted of one actual course, a version of an actual course altered to reflect best practices criteria as identified by the IDPs, and a version altered to be deficient in relation to the best practices criteria. Experts were asked to assure that these three courses were on a continuum, with a best practices course on one end, a course that does not reflect best practices on the other end, and the actual course in the middle.

2.2.3.3 Calibration Team Meeting

The calibration teams were convened at a two-day meeting in Orlando, Florida, to review and use the results of the online scoring of the three courses per subject area and to refine the instrument based on the results of this exercise. An 80% consensus rule was applied for all changes. Each calibration team made changes to the content and structure of the IDP instrument in its subject area.

As a result of the meeting, PE language, PE importance ratings, and PE organization and structure were changed. For example, a PE that addressed the importance of ensuring equal access and opportunity to all students was added to each subject.

2.2.3.4 Weightings

Weights were applied at multiple levels of the instrument to be used in the scoring and identification process of best practices courses. Weightings were used to ensure that not all aspects of the instrument were considered to be equal. This was one of the key goals of the study—to determine which content, habits of mind, and instructional practices were more central or important to instructors of best practices college courses. The purpose was to be able to provide high school teachers with better information about the areas in which they should focus instruction. The weightings of content PEs in particular also assists test developers with determining which topics should be assessed on AP exams and which might be omitted or given less importance. Therefore, the calibration teams were asked to identify the relative importance and weighting of the PEs. They reviewed and modified the importance ratings for each. (See Appendix D - lists performance expectations and their respective importance weightings.) Calibration teams decided on the weights to be applied to the three sections (Topical Frameworks, Habits of Mind, Instructional Practices), as well as between instructor ratings and external rater ratings. Weightings were achieved by dividing 100% among the three sections and then again between the two ratings sources (instructor or external rater) to indicate relative importance. (See Appendix E - shows section-level and data source-level weighting matrices used in scoring courses.)

2.3 Data Collection

The data collection process consisted of several steps. First, institutions that received significant numbers of AP scores in the subject areas being studied were identified for recruitment. Second, chief academic officers or provosts of these institutions were invited to participate in the study by nominating courses and instructors. Those agreeing to participate were asked to appoint an institutional liaison charged with working with departments to nominate either best practices courses taught by instructors at the institution or best practices instructors. If an instructor was nominated, the instructor chose which course(s) he/she would rate, based on criteria we provided. Third, participating instructors submitting their courses were asked to rate their courses against the instrument developed for their subject area by indicating where evidence of each relevant PE could be located within their course documents. Fourth, participating instructors were asked to submit course documents, in particular a course syllabus, for further analysis. Fifth, specially trained external raters, who were also college instructors in the subject area, rated the submitted documents to determine the points on which they aligned with the instrument. Finally, as previously mentioned, a course scoring process was followed in which a series of weights was applied to both sets of ratings for each course—those by the instructor and those by external raters—before being aggregated into an overall score.

2.3.1 Institutional Recruitment

The College Board Advanced Placement Best Practices Courses Study called for participation by postsecondary institutions drawn from among those receiving the most AP score reports in each subject area. AP staff provided researchers with a list of the top 500 postsecondary institutions requesting AP score reports for each of the seven selected subject areas based on the absolute and relative number of AP score reports sent for the enrolled freshman class entering fall 2004. This was done to ensure including institutions for which students submitting AP test scores constituted a significant percentage of the entering class. In this fashion, both larger and smaller institutions were included in the sample.

2.3.1.1 Sampling methodology

A sampling purposive methodology was used to draw institutions from the top 500 colleges and universities receiving the highest number of AP score reports. Samples were drawn, beginning with the top 100 AP score receivers in each subject area, then progressing down the list in groups of 100 until recruiting goals for each subject area were achieved. When necessary, additional institutions were invited to help guarantee adequate representation across all Carnegie classifications. The goal was to ensure that a wide range of postsecondary institutions from among those receiving AP scores was invited to participate and that those receiving the most AP scores were most likely to participate.

Depending on the institution and the subject area, some institutions were invited in all subject areas, while some were invited in fewer areas. For example, a doctoral/research university may have been invited in all subject areas, whereas a specialized institution may have been invited only in its area(s) of specialization. The sample consisted of 889 institutions. Of the 889 invited, 234 agreed to participate. An additional attempt to enhance the diversity of the sample was made by inviting all Minority Serving Institutions (MSIs) including those that were not included in the top 100 score reports receivers. The breakdown for MSIs is presented in Table 1.

Table 1: MSIs invited to participate

MSI Type	In Top 100	Not in Top 100
Historically Black Colleges and Universities (HBCU)	22	24
Hispanic Serving Institutions (HSI)	53	50
Asian	34	1
Native American	0	9
HSI + Asian	4	0
Total	**113**	**84**

2.3.2 Courses and Instructors Nominated

Once institutions agreed to participate, provosts or chief academic officers of institutions in the sample selected a liaison to work with researchers. In addition to the nominated faculty, College Board AP

Readers and calibration team members were also invited to rate a relevant course they had taught in the past year.

2.3.2.1 *Institutional Liaisons Selected*

Liaisons were generally central administration officers who then contacted department heads and solicited nominations of up to four courses that they felt best exemplified the study's definition of best practices. A school was never excluded because it was unable to generate enough courses in each subject area for which it accepted an invitation to participate. This was done to ensure participation by institutions of all sizes.

2.3.2.2 *AP Readers*

Institutions were recruited with the help of individuals who had previously served as scorers, known as AP Readers, of the constructed response portion of AP exams. Researchers invited AP Readers to participate and, in some cases, to help recruit their own departments and institutions to participate. When AP Readers agreed to participate, researchers then contacted the department chair and the provost or chief academic officer of the AP Reader's institution to encourage nomination of additional courses, either in the AP Reader's subject area or in another subject area.

2.3.2.3 *Expert Nominations*

The final source of nominated courses was calibration team members. They were provided a list of participating institutions and asked to identify institutions that 1) they believed had best practices courses in their subject area, 2) were not yet participating, and 3) had not been invited to participate. Team members were also given the option of nominating individuals whom they had reason to believe were instructors of best practices courses.

2.3.3 Courses Rated by Instructors

Instructors were contacted via email and invited to participate by rating their courses using the online instrument. Throughout the data collection phase, instructors who had not completed their online course ratings were emailed reminders that included the URL and login information. Approximately one month prior to freezing the data for scoring, researchers contacted faculty who had not completed their online course ratings with a phone call reminder.

All participants completed consent forms, and researchers maintained confidentiality throughout the study and reporting phases. In addition, instructors of courses identified as best practices reviewed by the Course Validation Panel (presented in this report) or annotated for presentation to the AP commissions were contacted and informed of these intended uses and permitted to withdraw their course at any point in the process.

2.3.3.1 Measures and Procedures

Participating instructors rated their own courses in relation to the best practices criteria presented in the online instrument. The following characteristics were used as the definition of best practices to enable institutional liaisons to identify the most appropriate courses for study:

- Course focuses on content that is most important for success in sequent courses in the discipline and are generally considered as most appropriate by disciplinary organizations in the subject area being taught.
- Course develops important habits of the mind, such as critical thinking, analytic thinking, and inquisitiveness.
- Course helps students learn to understand the structure of knowledge in the discipline and to think like scholars in that discipline.
- Course is taught using pedagogically appropriate strategies to maximize student learning.

Courses were grouped into two categories—corresponding and pathway. The following definitions were used to help institutional liaisons and department heads distinguish between the two course types.

- **Corresponding courses (taught at entry level).** The corresponding course refers to the first non-remedial course that a student takes for college credit or for which a student receives college credit based on an AP test score. These courses directly correspond with or are believed to most likely reflect what an institution would expect a high school AP course to cover.
- **Pathway courses.** These are courses that an academic department would reasonably expect a student interested in pursuing a given subject area to enroll in following receipt of credit for the corresponding course. Pathway courses are often taken after the freshman year. Some were next in a subject's sequence; others were related to or built upon the corresponding course, but were not necessarily strictly sequential.

This study focused primarily on corresponding courses because they relate most directly to AP courses. Data were collected and analyzed for pathway courses in order to help reach more informed conclusions regarding the most appropriate expectations for the challenge level and important content that AP courses should contain so that students are prepared for the courses they are most likely to encounter in college if they receive credit for an AP course in high school. Of the 234 institutions agreeing to participate, 171 (73%) actually submitted course data, for a total of 770 courses. Table 2 shows the number of courses submitted per subject area.

Table 2: Number of courses submitted for each subject area

Subject Area	Number of Courses Submitted
Biology	149
Chemistry	166
Physics	139
Environmental Science	53
European History	73
US History	133
World History	57

Submitted courses ranged in length from quarter to full-year. Instructors nominated for more than one course in a sequence were asked to complete one instrument for the entire sequence. Of the 770 courses, 76% were semester-long and 18% were full-year. The remaining 6% were either one or two quarter in length. Similarly, two-thirds of the submitted courses (67%) were corresponding. (See Appendix F - shows breakdown of course length and sequence type.)

The final makeup of participating institutions is presented in Table 3, followed by sample descriptions for both of the data sources (instructor ratings and external rater ratings) in each subject area.

Table 3: Participating institutions by Carnegie Classification type for each subject area

Carnegie Institutional Classification	Bio.	Chem.	Physics	Envi. Sci	Euro History	US History	World History	Total*
Doctoral/Research Universities—Extensive	22	25	24	13	17	25	10	**41**
Doctoral/Research Universities—Intensive	15	16	12	7	7	17	7	**29**
Master's Colleges and Universities I	22	22	15	7	15	29	16	**59**
Master's Colleges and Universities II	3	0	0	0	2	0	1	**4**
Baccalaureate Colleges — Liberal Arts	8	6	9	6	6	6	4	**18**
Baccalaureate Colleges—General	1	4	0	1	2	1	2	**7**
Baccalaureate/ Associate's Colleges	0	0	1	1	1	0	1	**1**
Associate's Colleges	1	3	2	1	0	4	1	**8**
Specialized Institutions—Other specialized institutions	0	0	1	1	0	0	0	**2**
Specialized Institutions—Schools of art, music, design	0	0	0	1	0	0	0	**1**
Specialized Institutions — Schools of engineering and technology	0	0	1	0	0	0	0	**1**
Total Institutions that Submitted Data*	**72**	**76**	**65**	**38**	**50**	**82**	**42**	**171**

** Total does not equal sum across all rows as most schools are participating in more than one subject area.*

2.3.3.2 Measures and Scales

Instructors rated their courses online using the instrument developed through the process described previously. Before beginning the online instrument, instructors were asked to create a course profile that contained information such as: course description, prerequisites, policies on plagiarism and grading, presence/absence of discussion sections and labs, use of online course management system to be used during the course document alignment ratings.

Instructors then completed each of the three sections of the instrument: Topical Framework (content knowledge), Habits of Mind, and Instructional Practices. Each contained a series of specific statements, called Performance Expectations (PEs), that detailed potential practices in each of the three sections. Instructors were asked to select for each PE its importance in the course they were rating. Instructors chose one of the following for each PE: Most Important, More Important, Less Important, Least Important, or N/A if the PE was not taught in the class.

2.3.4 Course Documents Submitted by Instructors

After completing the course rating, instructors submitted a syllabus and other course documents, such as tests, assignments, and labs and labeled each document by type. Researchers used these materials during the course document alignment ratings. Each document was converted to a PDF that was automatically linked to the submitting instructor's course and instructor rating. Personal identifying information was removed to the degree practical by the instructor or by the researchers. Instructors also had the option of mailing or emailing their documents independent of the online instructor rating process.

2.3.5 Course Documents Rated by External Raters

Of the 770 courses submitted, 568 (74%) met the minimum criterion of providing at least one acceptable document. Two external raters then rated the course. Table 4 shows the number of externally rated courses per subject area.

Table 4: Number of courses with documents for each subject area

Subject Area	Number of Rated Courses
Biology	85
Chemistry	114
Physics	117
Environmental Science	29
European History	58
US History	113
World History	52

2.3.5.1 Training and Reliability Procedures

As noted previously, instructors submitted a syllabus and other course documents, such as tests, assignments, and labs. External raters then rated these documents in relation to the same PEs that instructors used to rate their courses.

Rater Selection

Researchers selected individuals who submitted an application through an online process to become an external rater. They were chosen from the following three groups: 1) document raters who had participated in previous CEPR/EPIC projects, 2) individuals who had previously served as AP readers, and 3) individuals who had been nominated for the course validation panels but were not selected or declined participation. Of the nearly 300 potential raters, 25 candidates per subject area were identified as meeting all criteria required to participate in the document rater training. They had one or more of the following qualifications: high recommendations as a College Board Validity Study document rater, calibration team member, instrument development panel member, or course validation panel nominee.

A chief rater was hired in each of the seven subject areas to develop content to be used in training raters and in reliability-checking procedures. Chief raters were individuals who had been successful document raters for previous CEPR projects or had served in another capacity for the current best practices study.

Chief raters possessed at least two of the following three attributes:

- Five years of experience scoring against standards in their particular subject area (state, College Board or otherwise)
- In depth knowledge of AP course in a specific subject area
- Capable of breaking down content into pieces and analyzing those pieces to create benchmarks, develop a scoring guide, and provide feedback

Training and Certification to Rate Course Documents

Nine chief raters spanning the seven subject areas set criteria used to train raters and to perform reliability checks during the rating process. Chief raters chose seven courses from among all courses submitted by instructors in each subject area. Selected courses contained at least one document that was both comprehensive and a clear illustration of what was being taught or assessed. In other words, each course was appropriate as a means to establish a criterion-based judgment system that could then be applied to other documents or could be used to assess an external rater's ability to apply the criterion-based judgment system. Chief raters rated the seven courses, decided upon correct rating answers, and provided a rationale for each rating.

The training module consisted of three courses that had been prepared by the chief raters. Raters scoring 60% or above on the first training course proceeded to the second training course. Raters scoring less than 60% on the first and/or second training course saw the ratings and reasoning given by the chief rater and were encouraged to contact the chief rater with any questions regarding discrepancies between the ratings. External rater candidates received certification to rate course documents by achieving a minimum of 60% on two out of the three training courses in the training module. Given the complexity of the rating task, chief raters were allowed to interview raters who did not pass the training exercise and make an informed judgment of their ability. This option was only utilized for raters with exceptional expertise as identified from their CVP nomination or their experience as an AP Reader. Table 5 presents the number of external raters who passed the training for each subject area.

Table 5: Number of external raters who passed the training, by subject area

Subject Area	Number of external raters
Biology	12
Chemistry	9
Physics	11
Environmental Science	9
European History	12
US History	14
World History	10

Course Document Rating and Reliability Checks

Each course was independently rated by two trained external raters based on the content of all submitted course documents. Individual external raters rated between 1 and 90 courses.

Reliability checks required raters to meet the initial qualification criteria once again. This was accomplished through the use of benchmark courses, specially prepared courses that were inserted at particular intervals, unknown to the rater. The first benchmark course was assigned as the sixth course for rating, allowing each rater to have rated only five courses between certification and the first reliability check. The second benchmark course was assigned as the 16th course.

Raters scoring less than 60% on any benchmark course were first allowed to view the correct ratings and rationale given by the chief rater and were encouraged to call the chief rater with any questions regarding discrepancies between the ratings. These raters were then assigned another benchmark document to score before being allowed to resume rating. They were only allowed to resume rating if they successfully scored this second course at the qualification level.

Inter-rater reliability remained high throughout the rating process. Agreement among raters was consistent at 83% or higher in each of the seven subject areas. Table 6 illustrates the overall percentage of agreement among raters within each subject area.

Table 6: Overall percentage of agreement among raters

Subject Area	Inter-rater Agreement
Biology	86%
Chemistry	85%
Physics	87%
Environmental Science	83%
European History	90%
US History	88%
World History	87%

2.3.5.2 Documents Analyzed

External raters analyzed the course profile and documents submitted by the instructor. Documents most commonly consisted of the course syllabus, exams, and assignments. These documents were rated using as the reference point the same PEs the instructors used. However, raters used a dichotomous scale of "Evident" or "Not Evident" because it was not reasonable to expect raters to discern the relative importance of each element present in a course. Raters made their determinations by applying the following criteria:

Evident

- Explicit presence: The PE was explicitly or concretely evident in the document.
- Implied presence: Sufficient implication that the PE was evident in the document even though there was no concrete evidence.
- Inferred presence: A preponderance of instructors would cover this PE given the explicit evidence of other related topics in the document.

Not Evident

- No evidence, concrete or otherwise, that the PE was in the document.
- Insufficient evidence in the document to imply coverage even though the general topic of the PE was mentioned.
- Insufficient evidence in the document that the topic was covered to the depth implied by the PE.

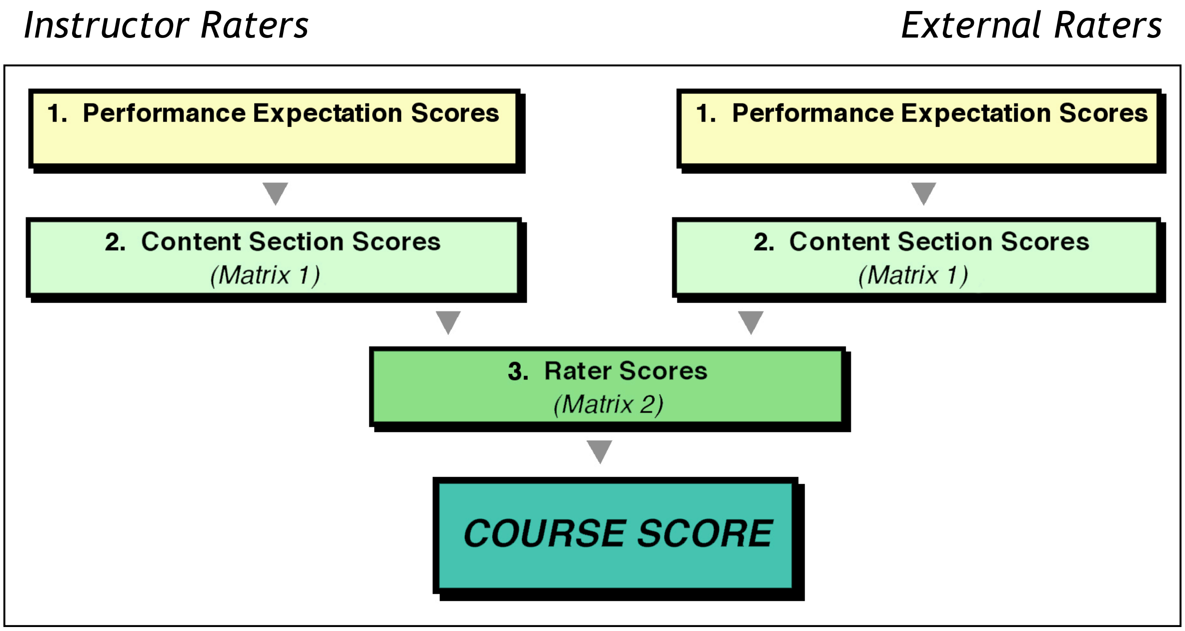
2.3.6 Course Scoring

As noted previously, weights were applied to instructor and external ratings for each course in order to compute an overall score. Figure 2 shows this aggregation process that resulted in a course score.

Each step designates a level at which weights were applied to the score. Score totals were added across each of these steps. Calibration team members collectively determined the weights to be applied, tailoring all but the first weighting matrix to each subject area. (See Appendix E - shows section-level and data source-level weighting matrices used in scoring courses.)

In the first step, the PEs were weighted in such a way as to give the most "credit" to items that met three criteria: 1) were rated Most Important in the instrument, 2) were rated Most Important by the instructor, and 3) were rated Evident by the external rater. Conversely, negative points were given for items rated Most Important by the instrument and then rated by the instructor as Less/Least Important or N/A or as Not Evident by the external rater.

Figure 2: Course-scoring process


Instructor Raters
External Raters
1. Performance Expectation Scores
1. Performance Expectation Scores
2. Content Section Scores (Matrix 1)
2. Content Section Scores (Matrix 1)
3. Rater Scores (Matrix 2)
COURSE SCORE

The weighted PE scores were summed within each section (i.e., topical framework, habits of mind, and instructional practices) in the second step in the diagram. These sections were then weighted as determined by the calibration teams, with different weights applied for the instructor rater than for the external rater. In the fourth step of the diagram, the weighted section scores were added together to form a course score comprised of instructor and external rater scores. These course scores were weighted one last time to give more or less weight to instructors or external raters and then added together to determine an overall course score.

Because the various weightings differed among the subject areas, the numeric value and range of scores differed as well. This means that absolute scores cannot be compared across subject areas. Despite differences in the weighting patterns followed in each subject area, course scores in all subject areas resulted in a relatively normal distribution pattern within each subject area. This is taken to indicate instrument sensitivity to differences among courses and as a confirmation that the process resulted in a subset of courses in each subject area that reflected best the criteria identified as best practices for that subject area. Figures 3 – 9 present the distribution of course scores as histograms for each subject area.

Figure 3: Distribution of course scores for biology

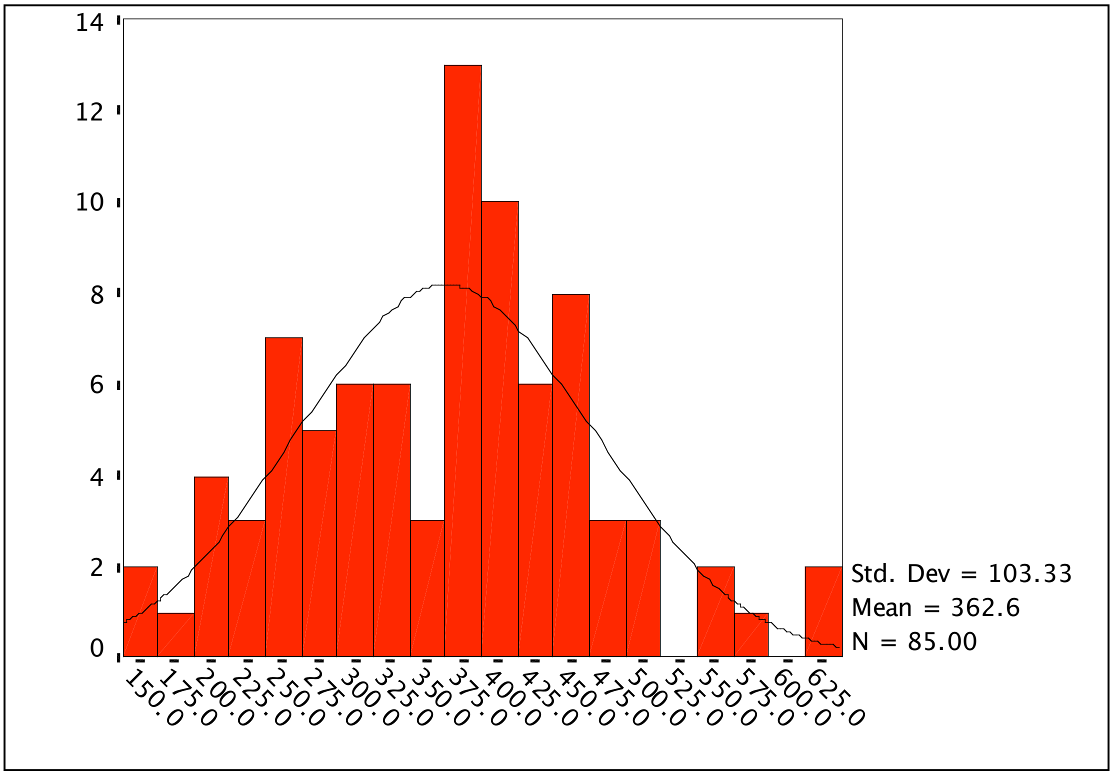


Figure 4: Distribution of course scores for chemistry

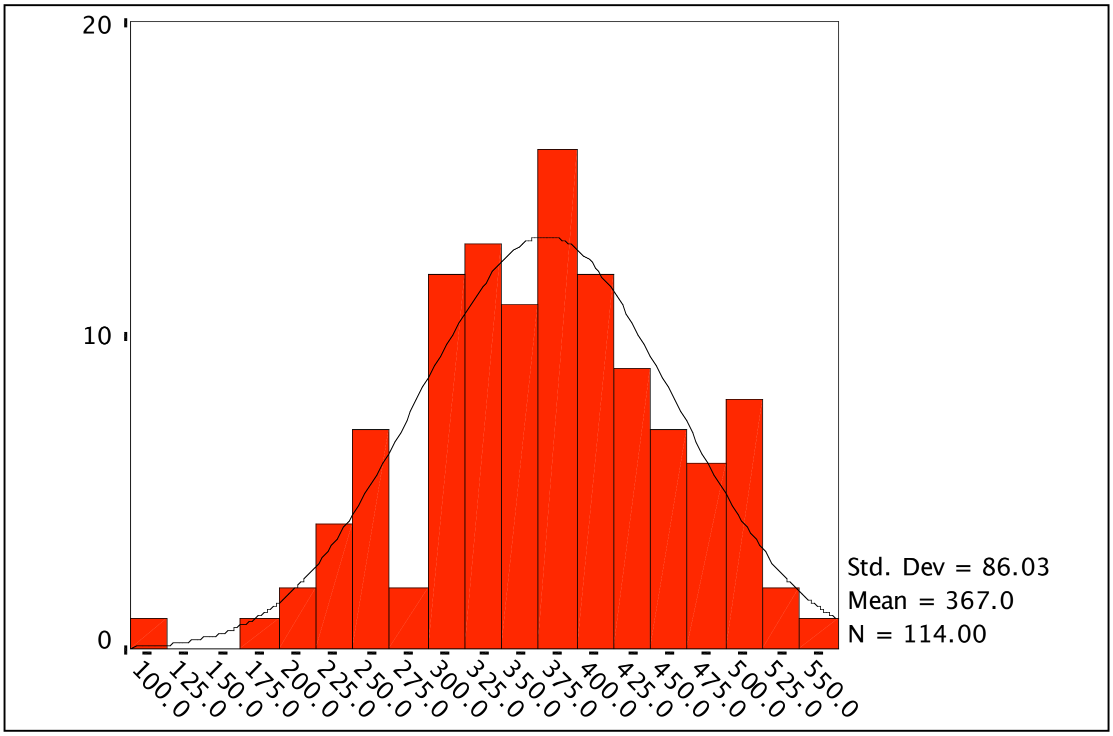


Figure 5: Distribution of course scores for physics

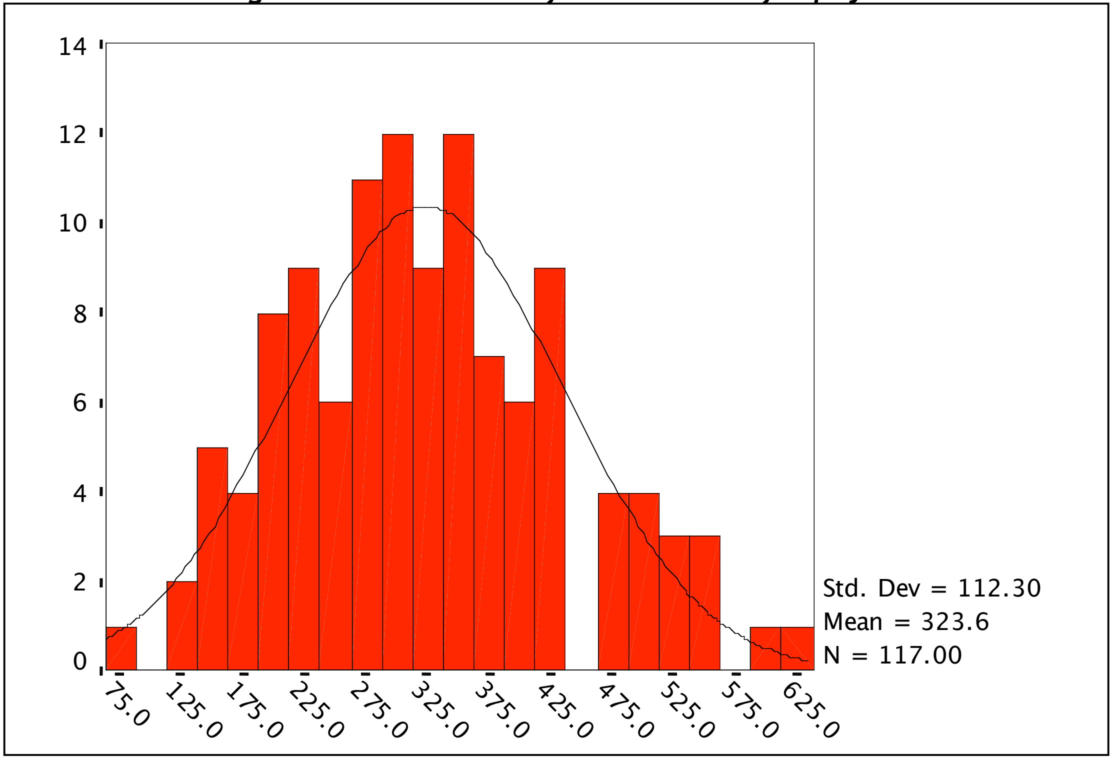


Figure 6: Distribution of course scores for environmental science

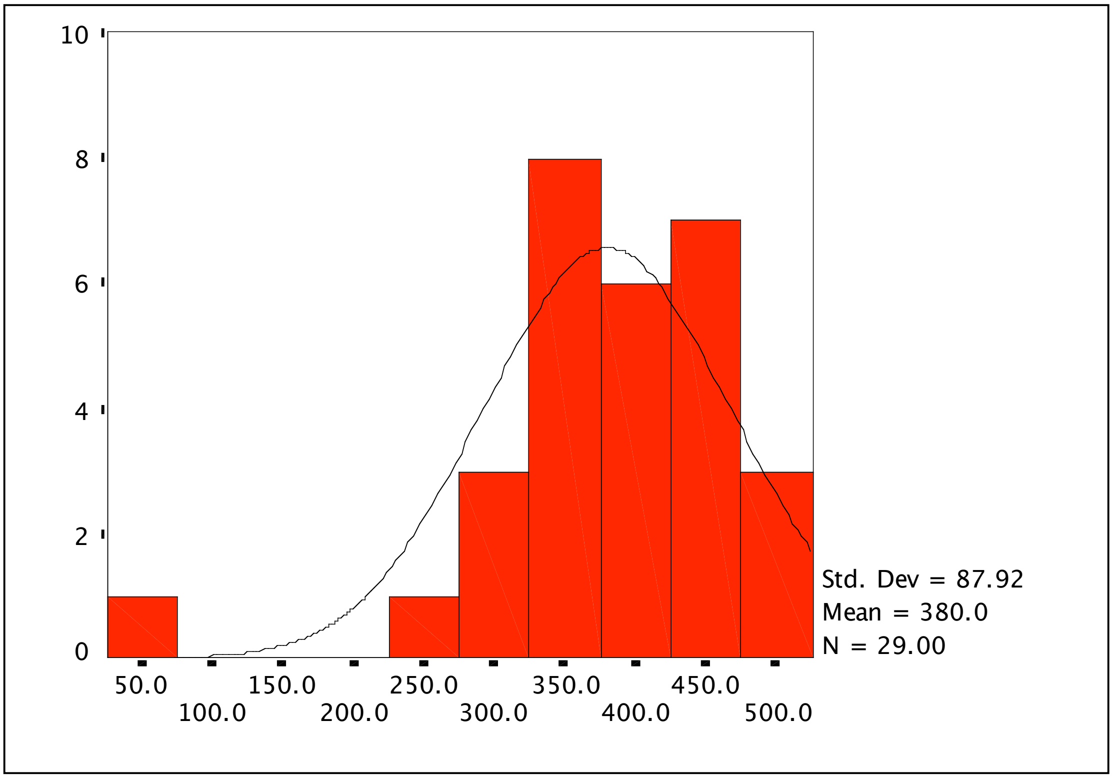


Figure 7: Distribution of course scores for European history

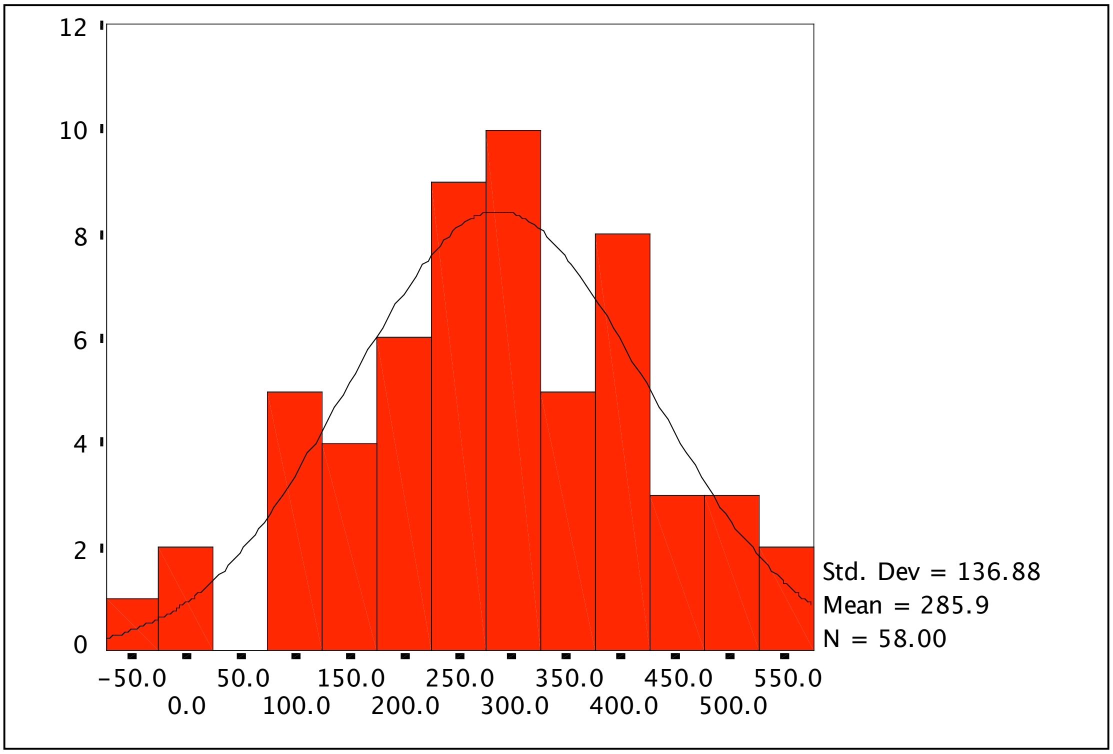


Figure 8: Distribution of course scores for US history

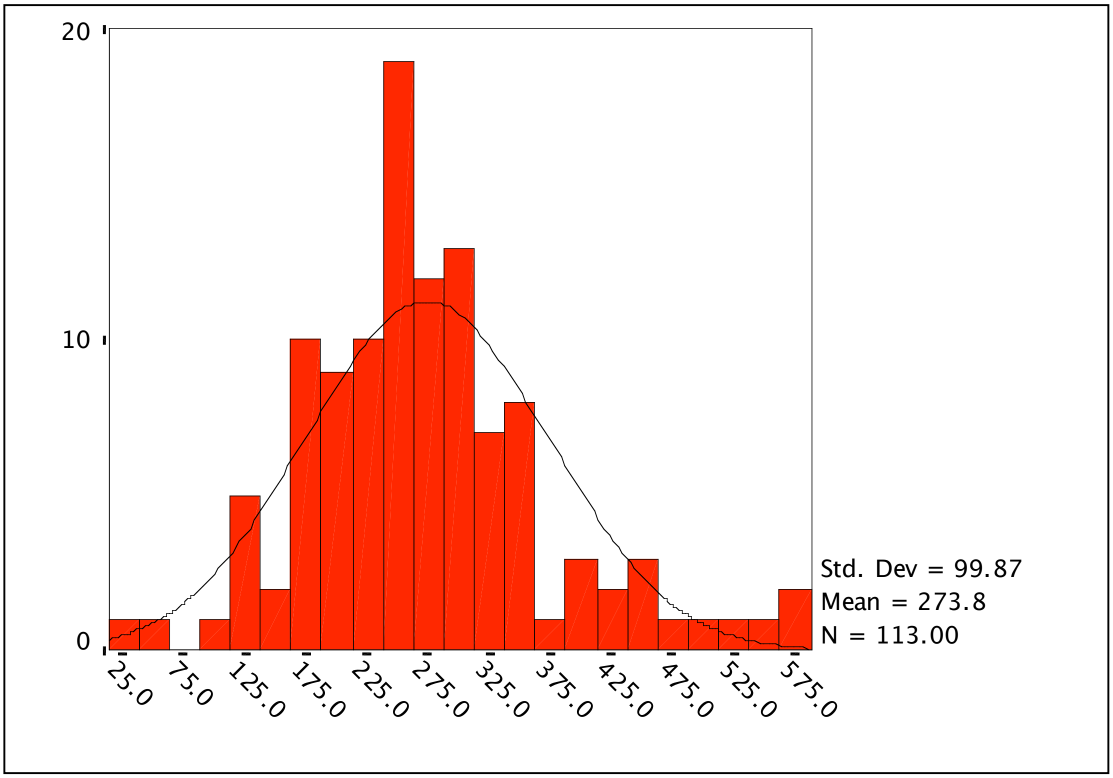
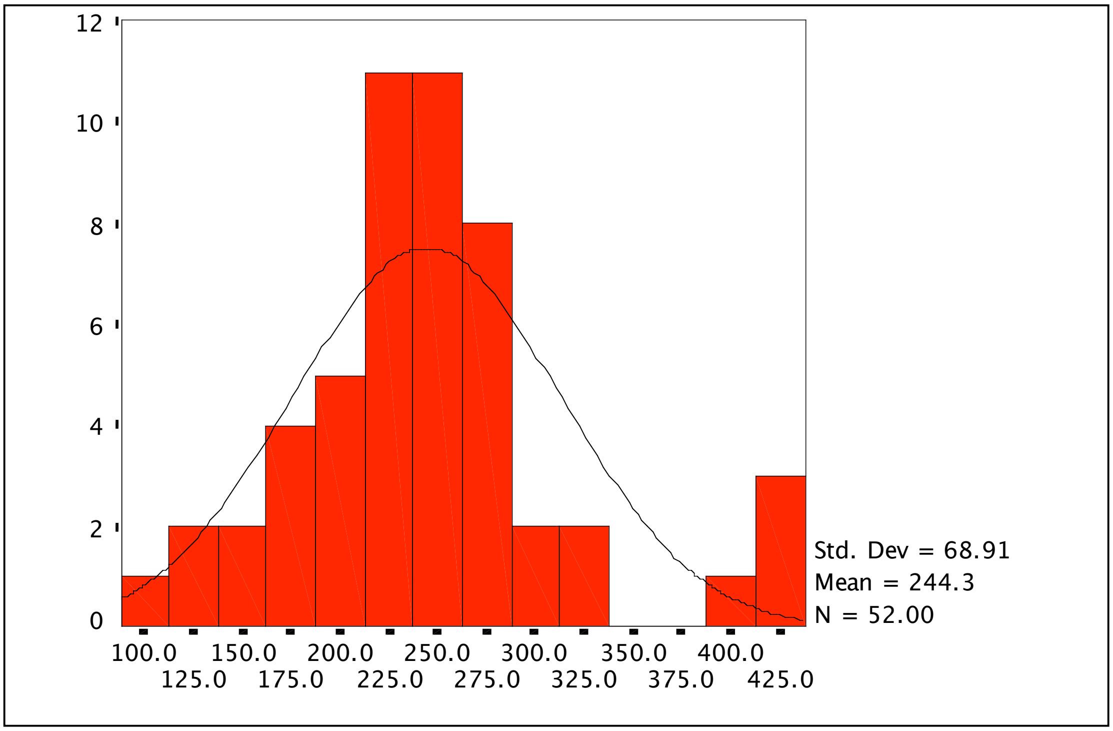


Figure 9: Distribution of course scores for world history



Ratings by instructors and external raters were compared to identify the percentage of agreement between the two groups. In order to do this, the five-point importance rating scale used by instructors (i.e., Most Important, Least Important, N/A, etc.) was converted to a two-point scale to be comparable with the dichotomous evidence scale (i.e., Evident/Not Evident) used by raters. For this rating scale conversion, all levels of importance were recoded as "Evident" and any rating of N/A was recoded as "Not Evident." The comparison yielded a moderate agreement rate between instructors and raters, ranging from 57% to 75% across subject areas. This less-than-complete agreement reflects the fact that instructors were able to identify elements as present in their courses that were not evident to external raters based on the documents submitted. Table 7 illustrates the overall percentage of agreement between instructors and raters within each subject area.

Table 7: Overall percentage of agreement between instructors and external raters

Subject Area	Inter-rater Agreement
Biology	57%
Chemistry	57%
Physics	59%
Environmental Science	70%
European History	75%
US History	68%
World History	75%

2.4 Best Practices Course Validation

Seven Course Validation Panels (CVPs), one in each subject area, were convened to review best practices courses that received the highest scores from the rating process. The purpose of each CVP was to determine if the courses that received the highest scores were in fact best practices courses in the subject area. This method of expert validation provided an additional check or screen to ensure that selected courses were highly representative of best practices in each subject area and, thus, models for high school instruction.

2.4.1 CVP Recruitment

Panels consisted of experts nominated by national disciplinary associations and by associations concerned with undergraduate education, particularly those with a focus on improving student learning through instructional practices, course structure, and cultivating habits of mind. Researchers and AP staff members identified key organizations in each subject area. Outreach to these organizations focused on soliciting each organization in order to identify leading subject area experts to serve on panels and on helping each organization understand the study's methods and goals. Two meetings were held in Washington, DC, and all solicited organizations sent delegates to one of these meetings where they were thoroughly briefed on the study. When familiar with the study, the organizations' delegates were asked to nominate at least three individuals to serve on the CVPs. Nominees were invited in the rank order identified by each organization. (See Appendix G - lists Course Validations Panelists' institutions and associations.)

2.4.2 Course Validation Process

2.4.2.1 Pre-CVP Work

Subsequent to nomination by a national organization, course validation panelists participated in a two-part orientation and training process. First, they reviewed the study's methodology and the best practices criteria in the instrument. Second, they participated in an orientation phone conference prior to conducting any validation work. This orientation process was designed to assure a common understanding of the study's methodology and to identify and address questions prior to the CVP meeting. Anticipating methodology-related questions helped to ensure that panelists would be applying similar or identical criteria and methods as they sought to endorse best practices courses and identify best practices course characteristics.

Following the orientation process, panelists were supplied with a URL that was linked to the highest overall scoring courses. Panelists examined these courses and their associated material to identify those that merited discussion during the CVP meeting. Course materials included the instructor ratings, the external rater ratings, and the course portfolio consisting of all course artifacts submitted by the instructor.

Panelists conducted global reviews of the highest scoring courses by reviewing course materials in relation to the instrument in each subject area. They were further instructed that the courses did **not** have to be best practices in every aspect but should show attributes that might be considered best practices. The following definitions were provided to help panelists determine the basis upon which a course should be advanced to the CVP meeting for further review:

- **Advance.** This course merits further discussion with my colleagues; it shows possible attributes of a best practices course.

- **Do Not Advance.** This course shows no possible attributes of a best practices course and does not deserve further consideration.

Panelists provided a rationale for each course they recommended for advancement. In addition, they identified supplemental information they would like to see gathered for each course and presented at the CVP meeting. Instructors of the candidate best practices courses to be reviewed during the course validation panel meeting were contacted to supply supplemental information in the form of additional course artifacts and responses to a set of open-ended questions. (See Appendix H - lists the supplemental questions to instructors of top scoring courses that were reviewed by the CVPs.)

In each subject area, courses recommended for advancement by at least half of the panel were forwarded to the CVP meeting. The goal was to have 15 courses in each subject area reviewed at the CVP meeting. In some cases, researchers conducted a second course-scoring process in order to include data received after the first course scoring. This second scoring resulted in best practices courses being advanced to the CVPs if they had higher overall scores than those reviewed in the pre-CVP exercise **and** if half of the panel had not recommended 15 courses for advancement in that subject area. Table 8 presents the number of courses in each subject area advanced by panelists and the number advanced through the second scoring process courses.

Table 8: Courses advanced to CVP

Subject	Pre-CVP Advanced Courses	Second Scoring Advanced Courses
Biology	13	2
Chemistry	11	4
Physics	11	4
Environmental Science	12	3
European History	13	2
US History	15	0
World History	12	3
Total	**87**	**18**

2.4.2.2 CVP Meeting

Course validation panelists were convened at a one-day meeting in Arlington, Virginia, to review the candidate best practices courses. Each panelist presented one or more of these courses to the rest of the panelists in each subject area. The panelists presented the attributes of the course as they related to the best practices criteria in that subject area. The panel then discussed each course and, after extended and extensive discussion, voted to classify each course into one of three categories: “Best Practices Course,” “Contains Attributes of a Best Practices Course,” or “No Endorsement.” Two-thirds of panelists had to recommend a course to be best practices for it to receive that designation.

In addition, panelists specified the PEs that were most important in a best practices course. For all courses that received an endorsement, panelists also identified the course attributes as they related to the best practices criteria in that subject area.

3 Findings

3.1 Overview

This chapter summarizes the four major sets of findings from the study. These findings help define what constitutes a best practices course in each of seven subject areas. The chapter begins with a description of the instrument that was developed to identify the characteristics of best practices courses. The instrument itself is a valuable resource in its own right, given the extensive development process undertaken to develop it. The second source of information consists of the aggregate scores on the instrument for all courses in a subject area. These data provide a frame of reference within which the most highly rated courses can be viewed. How are they similar to and different from the overall set of courses that were rated by instructors and external raters using the instrument? Third are the exemplary courses themselves. These are the ones that scored the highest on the instrument, were scored highest by external raters, and were validated by the Course Validation Panel. To help guide the AP commission members in interpreting what it is that makes these courses truly best practices, they have been thoroughly annotated by experts in the field, all of whom participated in one or more aspect of the study's review processes (instrument development, calibration, rating, and course validation). The annotations connect the best practices criteria on the instrument with specific aspects of each course to provide an operational description of what the criteria mean in practice. Finally, the study yielded composite best practices courses. These are courses that draw from more than one best practices course to create one course that represents best practices throughout. This technique allows outstanding practices from multiple courses to be captured and showcased in a course that others could emulate if they chose to do so and that fully informs the AP commissions.

3.2 Final Version of the Instrument

The instrument employed to identify best practices courses is itself an important resource both for the commissions and, eventually, for test designers and item writers. (See Appendix I - shows the final data collection instrument.) The instrument was developed through an extremely thorough process of repeated expert reviews that is much more systematic than any similar effort previously undertaken in this area.

In addition to being a reference point for the analysis of best practices courses, each subject's instrument represents a set of empirically derived standards that clearly delineate what is most important for students to know and be able to do in an AP course in that subject area. This information will help the College Board address the issue of breadth versus depth in AP instruction and could legitimately be used in any subsequent test revision process, whether in conjunction with a commission's findings or as supplemental data. This clarity on what is most important can also be used to guide professional development for AP teachers and help them know the areas they must be certain to address in their courses and the areas where they have much greater discretion in their teaching.

Furthermore, the instrument also contains two categories in addition to content knowledge—habits of mind and instructional practices—that have not previously been chronicled with this specificity and detail. These two areas are identified consistently by higher education faculty as being as or more important than specific content knowledge. This additional information will, at the least, assist high school AP teachers in knowing how to organize their teaching. The habits of mind data can serve as a particularly powerful framework within which to influence the evolution of AP professional development—encouraging teachers to more fully develop skills, such as complex thinking, that will improve student transition to college. The instructional practices highlight the overlap between good college teaching and good high school teaching. Many instructors at both levels will readily recognize many of these practices as being common to their classes. This information may provide a basis for high school and college instructors in AP subject areas to trade more ideas and techniques, particularly those that address key instructional practices identified as most important to AP courses.

3.3 Aggregate Ratings of Best Practices Courses

The aggregate data that report frequencies for each item on the instrument serve as a useful supplement to the instrument itself and to the annotated best practices courses. This information stands midway between the general and undifferentiated instrument Performance Expectations (PEs) and the specific and highly contextualized best practices courses annotations. These results are reported in an unweighted fashion to present a picture of what might be described as normative practices within best practices courses in a subject area relate to the instrument PEs. In each subject area, the results from the scored courses are presented as mean scores of importance for each PE. (See Appendix J - shows performance expectation findings for all courses and Appendix K - for the legend for findings in appendix J.)

The fashion in which all the nominated courses map onto the PEs provides broader insight into current practices in the field and indicates the variance present even among the best courses. This information is useful because it helps create a more complete picture of best practices by capturing the range that exists among all the courses that institutions perceive to be best practices. This helps defuse some of the arguments about the one best way to teach a course and acknowledges a variety of approaches while simultaneously identifying the normative practices that do exist among best practices courses.

These data that are generated help identify PEs that emerge as most important across the range of best practices courses in a subject area and then again in the specific best practices courses that received the highest scores and were validated. This comparison process can create even greater confidence that a subset of PEs is absolutely important to success in college courses.

(See Appendix L - shows performance expectation findings for only those top scoring courses that were reviewed by the CVPs during the panel meetings and See Appendix M - legend for PE findings exemplary courses table used in appendix L.)

3.4 Exemplary Courses with Annotations

Exemplary courses were those that were endorsed by each CVP as “Best Practices Courses.” These select courses were further analyzed to identify the ways in which they put best practices criteria into practice. This was done through annotation, whereby a panelist from each subject area performed a line-by-line review of all of the course’s documents and any supplementary information to identify those elements of the course that demonstrated or corresponded to specific PEs in that subject area. Where possible, the panelist was someone who had also served on more than one of the following: the IDP, the calibration team, or the course validation panel in the subject area. These experts provided detailed rationale statements that explained why and how each selected course element demonstrated one or more PE. The result was a set of embedded annotations of critical course elements that further highlight what is important in each exemplar best practices course. (See Appendix N - shows performance expectation findings, including annotations for fully endorsed exemplary courses by the CVPs during the CVP meetings.)

Findings present each exemplary course portfolio in its entirety with rationale statements explaining why and how descriptive elements reflect one or more PEs. This set of courses offers a contextualized view of the PEs in practice. These annotated courses will be particularly useful to the commissions as a means to view how sometimes-abstract assumptions are operationalized in practice. (See Appendix O - annotated course portfolios for fully endorsed exemplary courses.)

3.5 Best Practices Composite Course

Experts who were selected through the same process just described also created a descriptive and detailed narrative of a “model” best practices course for each subject. These courses are composites whose elements are drawn from courses endorsed fully by the Course Validation Panels from courses the panels identified as having numerous attributes of a best practices course. These composite courses provide the opportunity to create the highest quality examples of courses that incorporate all desired elements into one consistent format. (See Appendix P - model composite course and model assessment materials.)