I. Introduction

The Policy and Program Studies Service (PPSS), Office of Planning, Evaluation and Policy Development, U.S. Department of Education (the Department), requests clearance for case study data collection activities for an evaluation of Evidence-based Practices in Online Learning. This evaluation is conducted under the authority of Title II, Part D of the Elementary and Secondary Education Act (ESEA) of 2001. Through a systematic and comprehensive review of the literature, analysis of K-12 student achievement data, interviews, and observations, the study will describe, using scientific research standards, what is known about the conditions and implementation practices that support effective online learning within both K-12 and teacher populations. In addition, this project will identify and describe examples of promising online learning practices and examine how these practices may vary on the continuum between face-to-face instruction and independent online learning by students. In this submission, we request clearance for the study design, sampling strategy, and case study data collection activities.

Distinctive aspects of this evaluation of online learning are the following:

- Conceptual framework based on prior research to guide the scope of case studies to be included, and to provide a common terminology for the project team, the Department, and external technical experts.
- Balance of breadth and depth in the research review, integrating disparate literature bases related to online learning practices and conditions from multiple fields (e.g. K-12, postsecondary, industry, teacher professional development, and military), while focusing in-depth analyses on studies with statistically rigorous research designs and quantitative measures of student learning outcomes.[1]
- Quantitative analysis of student outcome data for online and face-to-face K-12 students. Secondary data will be analyzed to compare student outcomes in face-to-face and online learning contexts for two cohorts of students after controlling for prior achievement. Two distinct populations will be studied. Students enrolled in the Florida Virtual School (FLVS) will be compared against their face-to-face counterparts, as will students enrolled in virtual courses offered by a large suburban school district.

[1] Inclusion/exclusion criteria for the meta-analysis are included in Appendix G.

- Site visits to gather qualitative data, including interviews with developers, administrators, instructors, and students, at both K-12 and TPD sites identified through the literature reviews and secondary data analyses described above. Researchers will verify and expand upon findings from earlier analyses, as well as observe the use of applications at both exemplary and typical sites. K-12 site visits will be conducted with FLVS developers, as well as with developers from a large, suburban school district. Up to ten FLVS implementation sites will also be visited. Comparing exemplary and typical sites for online learning applications will help isolate promising practices by focusing on those that appear predominantly in exemplary sites but not in typical sites. We will also provide rich descriptions of the practices associated with improved student learning, as identified in the meta-analysis and K-12 secondary data analysis, and of the conditions that appear to be necessary to successfully implement these practices.

This document supports the information provided in Form OMB-83-I. The remainder of this section provides background for the study and a conceptual framework for study activities, as well as evaluation questions and our approach to data collection. In the Supporting Statement for Paperwork Reduction Act Submission that follows, we provide a justification for the study and additional details regarding the collection of information employing statistical methods.

A. Background and Conceptual Framework

Modern distance education today includes online or "e-learning" offerings that range from conventional didactic lectures or textbook-like information delivered over the Web to Internet-based collaborative role playing in social simulations and highly interactive strategy games. Examples may include primary grade students working on beginning reading skills over the Internet, middle school students collaborating with practicing scientists in the design and conduct of research, and high school dropouts taking courses online to recover credits and qualify for graduation. In addition, entire school programs provided over the Internet are either offered or in the planning stage in many states. By fall 2007, 28 states had online virtual high school programs (Tucker, 2007). The teachers of K-12 students may also participate in online education, logging into online communities and reference centers and earning in-service professional development credit online. Appendix A provides a glossary of terms used commonly in relation to online learning.

A study of such diverse learning options requires a clear conceptual framework to guide efficient data collection. The characteristics of the conceptual framework are based on a preliminary review of the research literature, and we will use the conceptual framework to classify the types of online learning to be studied. There are three components to our framework: 1) whether the activity serves as a replacement or enhancement to conventional teaching; 2) pedagogical approach; and 3) synchronicity of instruction. Each of these is described in further detail below.

In our review of the research literature, one of the most basic characteristics for classifying online activities is whether the activity *replaces* face-to-face instruction (e.g., a virtual course) or *enhances* face-to-face instruction (i.e., online learning activities that are part of a course given face-to-face). This distinction is important because the two types of applications have different objectives. A successful replacement application provides equivalent curriculum to conventional instruction without sacrificing student learning outcomes. If student outcomes are identical online or face-to-face, then online instruction could be a more cost-effective option in settings where few students are situated in a particular geographic locale (e.g. rural students or specialized courses). In contrast, a successful enhancement application produces better learning outcomes than with face-to-face instruction alone. If the outcomes were identical, the enhancement application would be considered a waste of time and money.

A second dimension is the pedagogical approach of the activity, often referred to as "knowledge acquisition," which concerns who (or what) determines the way learners acquire knowledge. The contractor proposes to use the three-categories of *learning experiences* developed by Galvis, McIntyre, and Hsi (2006) which allows us to differentiate two types of student-centered learning experiences (active and interactive learning) and highlight the predominant characteristic of instructional activities (expository instruction). In each case, specific types of technologies can be associated with particular pedagogical approaches, as described below:

- Expository instruction— knowledge is transmitted using digital devices;
- Active learning—knowledge is built by the learner through inquiry-based manipulation of digital artifacts, such as simulators, games or microworlds; or
- Interactive learning—knowledge is built by the learner through inquiry-based collaborative interaction among learners; teachers become co-learners and facilitators.

Typically, in expository instruction, the technology delivers the content. In active learning, the technology allows students to manipulate digital artifacts in order to solve relevant questions; and in interactive learning, technology mediates human interaction either synchronously or asynchronously.

Many instructional offerings will blend these different forms of learning and also will combine online and face-to-face components. These "hybrids," so called because they integrate multiple approaches to instruction, require a careful examination of the particular program to make an informed judgment concerning the most appropriate classification.

Finally, a third characteristic important in categorizing online learning activities is the extent to which the activity is *synchronous*, with instruction occurring at a scheduled time in either a physical or virtual place, versus *asynchronous*, with a time lag between the presentation of instructional stimuli and student responses. Exhibit 1 illustrates the three critical dimensions in our framework for categorizing online learning offerings.

Exhibit 1. Conceptual Framework for Online Learning

Learning Experience Dimension	Synchronicity	Face-to-Face Alternative	Face-to-Face Enhancement
Expository	Synchronous	*Live, one-way web cast of online lecture course with limited learner control (e.g., students proceed through materials in set sequence)*	*Viewing web casts to supplement in-class learning activities*
	Asynchronous	*Math course taught through online video lectures that students can access on their own schedule*	*Online lectures on advanced topics made available as a resource for students in a conventional math class*
Active	Synchronous	*Learning how to troubleshoot a new type of computer system by consulting experts through live chat.*	*Chat with experts as the culminating activity for a curriculum unit on network administration*
	Asynchronous	*Social studies course taught entirely through Web quests exploring issues in U.S. history*	*Web quest offered as an enrichment activity for students completing their regular social studies assignments early*
Interactive	Synchronous	*Health care course entirely through an online, collaborative patient management simulation that multiple students work with at the same time*	*Supplementing a lecture-based course through a session spent with a collaborative online simulation used by small groups of students.*
	Asynchronous	*Professional development for science teachers through "threaded" discussions and message boards on topics identified by participants*	*Supplemental, threaded discussions for pre-service teachers participating in a face-to-face course on science methods*

Exhibit reads: Online learning applications can be characterized in terms of the kind of learning experience they provide, whether computer-mediated instruction is primarily synchronous or asynchronous, and whether they are intended as an alternative or a supplement to face-to-face instruction.

We will use the above three dimensions as a beginning framework to describe the implementation of online learning at both K-12 and TPD sites. These concepts will also be used to inform our instrument development so that our case study reports can adequately describe how instruction is delivered to K-12 students and teacher participants. Results of the meta-analysis and quantitative analysis will also be used to inform instrument development, so that we collect detailed information about the conditions and practices that were associated with student achievement in online learning

These three dimensions are not mutually exclusive; in each of them it is possible to "blend" the learning activity, combining any of the above-mentioned poles per dimension. And there are, of course, many other features of online learning, including the nature of the content to be learned (both subject area and type, such as fact, concept, procedure, or strategy) and the technology involved (audio/video streaming, Internet telephony, podcasting, chat, simulations, videoconferencing, shared graphical whiteboard, screen sharing, etc.). Many of these additional features will be picked up in our data collection instruments, and the data will be used either for descriptive purposes or to generate qualifications for best practices (e.g. a practice appears to work better for one subject or population than another). We have chosen to focus on the dimensions in our framework on the basis of our review of prior meta-analyses in distance learning. Bernard et al. (2004) found advantages for asynchronous over synchronous distance education, but Zhao et al. (2005), using a different study corpus, found that studies of distance learning applications that combined synchronous and asynchronous communication tended to have more positive effects than those of distance learning applications with just one of these interaction types.[2] Zhao et al. also found advantages for blended learning (which we term "face-to-face enhancement") over pure online learning experiences and of courses with more instructor involvement, as opposed to the more "canned" applications we have described as expository. Finally, we have included the setting dimension in our framework not only because instructional setting has been associated with effect size in some previous distance learning meta-analyses (e.g., Machtmes & Asher, 2000) but also as a pragmatic concern; as described below, different

[2] Both of these meta-analyses included video-based distance learning as well as Web-based learning and included studies where the outcome measure was student satisfaction, attitude, or other non-learning measures as well as those with a learning outcome. The meta-analysis forming the basis for case study selection in this study is restricted to an analysis of effect size for objective student learning measures in experimental and quasi-experimental studies of applications with Web-based components.

data collection approaches will be needed for applications that are used in non-classroom settings because users are not grouped in space or in time.

The above three characteristics capture some of the most important elements of online activities and together provide a manageable framework for differentiating among the broad array of online activities in practice today. With this approach we can meaningfully compare a variety of online approaches in order to portray the diversity of practice today as well as juxtapose one type, with its strengths and weaknesses, to another type.

B. Goals of the Evaluation and Evaluation Questions

This study will identify characteristics of online learning that are associated with student achievement through a meta-analysis and literature review of K-12 and TPD online learning studies, as well as through quantitative analysis of two K-12 student datasets: one from Florida and another from a large suburban school district. These analyses will be conducted independently (i.e. no attempts will be made to do pooled analysis using both case study quantitative datasets). Qualitative case study data will supplement these quantitative findings to highlight emerging promising practices in locations with established online learning programs, provide guidance for educational policymakers about how best to implement online learning, and establish a base for future qualitative and quantitative research. The proposed focus of the study is on the capabilities and potential of online learning for both K-12 education and teacher professional development. This research is necessary as online learning becomes increasingly popular in the education space, and as the perception that technology holds great promise for the improvement of K-12 and TPD education develops. Recent studies suggest that there may be more than one million students enrolled in online research in recent years, yet there is very little rigorous research to establish when and how online learning might be an appropriate intervention. This research will build on existing distance education research, which often does not include advanced Internet technologies and focuses on military and other training domains.

Our meta-analysis addresses online learning research, no matter what subject matter or content domain is covered by a particular study. This has not been done before. Prior meta-analyses, though potentially informative, did not focus on online technologies in K-12 or teacher professional development domains. Many distance education studies often focus on military and corporate training and higher education. In addition, some researchers have suggested that studies of online and distance learning are of questionable quality (Institute of Higher Education

Policy, 1999; Blomeyer, 2006), and that the good studies are fragmented across multiple literature bases (e.g., higher education, military training, etc). We have chosen an approach that integrates these seemingly disparate literatures in order find enough studies that will allow us to hone in on those practices that are associated with positive student learning outcomes. The main finding from the literature review was that few rigorous research studies of the effectiveness of online learning for K-12 students have been published. The meta-analysis of 51 study effects found that learning outcomes for students taking courses online exceeded those of face-to-face students, and that blended instruction is more effective than either purely online or purely face-to-face instruction. We also found that out of thirteen implementation variables, only the use of a blended approach and "time on task" emerged with a statistically significant positive influence on effectiveness. Online learning appeared to be an effective option for both undergraduates and graduate students and professionals; however, though positive, the mean effect size was not significant for six contrasts involving secondary and middle school subjects. Finally, the contractor examined characteristics of the studies to determine whether features of the studies' methodologies could account for obtained effects, but only "equivalence of curriculum and instruction" emerged as a significant moderator variable. Studies in which analysts judged the curriculum and instruction to be identical or almost identical in online and face-to-face conditions had smaller effects than those studies in which instruction between the two conditions varied.

Our case study work will be guided by these findings. Two cases have been identified based on the availability of sufficient quantitative data. Sites within each of these cases will be visisted. K-12 site selection will be guided by findings from the quantitative analysis of online and comparison-group student data from Florida and a large suburban school district. TPD site selection will be guided by the findings from a review of research focusing on how teachers are prepared to teach online. By grounding case study selection in the findings resulting from these analyses, the contractor will be able to hone in on those practices that have evidence of effectiveness. We will also look to the broader literature on distance education, even if online technologies are not used. We believe that many issues dealt with in earlier distance education applications and technologies are likely to remain relevant when new technologies or different populations are involved. The value of our work is that we will apply these findings to K-12 environments that use online technologies, and to teacher professional development environments that prepare teachers to teach online. Although a few good, related studies exist,

much is still not known about the conditions and practices in which online technologies can be used effectively with K-12 students and teachers.

Therefore, the goal of this study is to provide research-based guidance to policymakers, administrators, and educators on two key topics: how to optimally implement online learning for K-12 education, and how to prepare teachers to become effective online instructors. Fulfilling this goal requires a clear definition of the study scope and focus. We propose to focus on those conditions and practices associated with both K-12 and teacher professional development programs that appear effective in supporting student learning with online activities. The distinction between conditions and practices is one that the contractor has explicated in our prior work on the implementation of educational technology (Agodini, Dynarski, Means, Murphy, & Rosenberg, 2005; Means, Murphy, Shear, Gorges, Hu, & Sussex, 2006). *Conditions* are features of the context within which the technology is implemented that predated the introduction of the innovation and are relatively impervious to change. These would include the demographic characteristics of the learners, teacher or instructor qualifications, state accountability systems, and so on.

In contrast, *practices* describe the way in which the technology is implemented. These are choices made as part of implementation (for example, whether or not to have an online course facilitator). In choosing *whether or where* to use distance education (e.g., to teach mathematics for high school students or to teach world language for elementary schoolers), it is important for a policymaker to understand the conditions under which online learning is more and less effective. In deciding *how* to implement distance education, it is important to understand the practices that research suggests will increase the innovation's effectiveness (e.g., community building among participants, facilitator's role, blending work and training).

Using the results of the meta-analysis, K-12 secondary data analyses, and TPD literature review as a guide, the contractor will conduct a series of case studies of applications embodying the features found to be associated with effectiveness in prior research. K-12 case studies will examine online offerings provided by FLVS and a large suburban school district. Both the K-12 and TPD case studies will document practices in detail and across multiple sites, highlighting those that appear most promising (on the basis of an association with stronger learning outcomes obtained from the K-12 secondary data analysis). The evaluation questions the contractor proposes for guiding the case studies are:

- What are the key design features of online learning applications that are congruent with the research base in K-12 education?
- What are the key design features of emerging online learning approaches in K-12 education?
- What user conditions and practices are associated with more successful implementations of online learning applications (replacements for FTF)?[3]
- What user conditions and practices are associated with more successful implementations of blended online learning applications (enhancements to FTF)?[5]
- What conditions challenge implementation fidelity of online learning?

The conditions and practices associated with online learning may differ based on the type of online education under study. For example, a short course that promotes student learning of simple facts may not require the same support as a course that aims to encourage conceptual understanding or changes in behavior, as is the case with many teacher preparation programs.

Our literature review has provided us with a baseline list of conditions and practices to be verified during case study research. For example, there is some suggestion that the ability of the learner to control learning media is important (Zhang, 2005; Zhang et al., 2006). Several studies suggest that manipulations that trigger learner activity (Gao & Lehman, 2003) or learner reflection and self monitoring of understanding (Chung, Chung, & Severance, 1999; Cook, Thompson, Thomas, Thomas, & Pankratz, 2006; Suh, 2006) enhance individual online learning outcomes. On the other hand, provision of simple multi-choice quizzes does not seem to be effective (Lewis, 2007; Maag, 2004; Stanley, 2006). Attempts to provide scaffolds for monitoring of their learning for groups of learners appear to be less successful than those used in the studies with individual learners (Choi, Land, & Turgeon, 2005; Hron, Hesse, Cress, & Giovis 2000), but the provision of rules for online group interaction appears helpful (Weinberger, Ertl, Fischer, & Mandl, 2005). However, because online learning is new and practices are evolving, we expect that the case study work may uncover promising practices not yet identified in the literature. The evaluation report developed through this study will highlight effective, research

[3] "Success" is defined primarily on the basis of learning outcomes and secondarily on the basis of usage level (the product of number of users and average hours of use).

based practices and replicable educational interventions for practitioners and policymakers at federal, state, and local levels.

C. Data Collection Approach

Sound data on the state of online learning in education are needed to inform educational policy and practice. Collecting high-quality data involves careful attention to each detail in the data collection process, from asking the right questions during interviews to ensuring that the data collected are handled accurately and ethically. We have planned a set of interrelated strands of data collection combining quantitative and qualitative methods as follows:

- Case studies of a purposive sample of up to 12 K-12 sites. The contractor will visit FLVS developers, FLVS implementation sites, and online developers associated with a large suburban school district.
- Case studies of a purposive sample of up to 12 online applications and/or programs which prepare teachers to teach online.

The K-12 case studies will consist of up to two developer site visits (one with FLVS developers and one with developers in the large school district), and up to ten district or school visits to FLVS sites identified as either "exemplary" or "typical" through the quantitative data analysis. The TPD case studies will consist of up to 12 developer site visits; administrator, instructor, and student participation may be observed virtually during these developer interviews.

The goals of the K-12 developer interviews and visits are to (1) become thoroughly familiar with the online offering and its *intended* implementation and (2) obtain information on exemplary and typical user sites. Where possible, we will collect any quantitative, objective data (such as usage measures or mean gain on pre- and post-assessments) that developers have available. These objective data, in combination with the developer's evaluation of implementation fidelity, will be used in conjunction with results from the analysis of student achievement data to select exemplary and typical K-12 FLVS implementation sites. Implementation sites within the large suburban school district will also be selected, pending approval from district officials. Typical sites are defined as those that have shown average, rather than exemplary, student gains. The comparison between these two types of sites will help isolate those practices associated with effective use. If practices appear equally in both types of sites, they are less likely to be linked with gains in student achievement. Therefore, we will focus on

those practices that appear predominantly in exemplary sites in forming recommendations for future research.

The goals of the TPD developer interviews are to (1) become familiar with the training that is provided to teachers who are preparing to become online instructors, and (2) determine which skills are important for online teachers to possess. As in the K-12 developer site visits, the contractor will attempt to collect any quantitative, objective data that exists during these developer interviews. Ultimately, the information gained from these 12 case studies will provide a baseline for future research on the conditions and practices necessary for teacher professional development to be effective, as well as the skills that a successful online teacher must possess.

Interviews with both K-12 and TPD developers will obtain information concerning the major features of the online offering or training, its intended implementation, and the availability of objective measures pointing to more and less effective user sites. During developer site visits, researchers will request system demonstrations which will provide a context for more detailed information gathering concerning intended system use. Researchers will ask for log-in information for particular applications so that they can objectively assess the features and capabilities of the online programs themselves.

K-12 implementations that occur in school settings will be observed at the school. Up to ten FLVS implementation sites (five typical sites and five exemplary sites) will be selected based upon the findings from the secondary data analysis. (The contractor may also visit implementation sites at a large, suburban school district if approval is granted and there is sufficient evidence of school-level variation). At school-based sites, the contractor expects to meet with program administrators, information technology specialists, between one and three instructors, and a focus group of between three and eight students who are currently using the application.

Implementation of the online program that does not occur within a formal school context (virtual high schools being one notable example) presents a particular challenge for site visits. Traveling to a particular physical site, where the contractor may see only one student using an offering (as might be the case for advanced placement) courses targeted to rural settings), would not be a cost effective use of project resources. Accordingly, for non-classroom applications the contractor proposes to conduct “virtual” site visits, where researchers log into a particular offering and observe the use of the offering over a period of time. By logging into the

application, the researcher can have the same perspective as a teacher or student using the application. Once logged in, researchers will observe student and instructor use of the offering by engaging in online discussion, observing chat room and discussion board interactions, observing online instruction, and collecting information concerning student's outputs and achievement whenever possible.

Instructor and administrator interviews associated with virtual site visits will be scheduled via telephone and email communications and will be conducted over the telephone. In some cases, the contractor will have already interviewed instructors and administrators during the developer site visit, but the contractor will also schedule supplemental instructor interviews associated with specific offerings observed during the virtual site visit. As an analog to the student focus groups that the contractor will conduct in school-based site visits, the contractor will invite students to participate in focus groups that will be conducted in a chat room or similar virtual space set up for the purposes of the study.

Collecting data from multiple types of respondents, including developers, instructors, and students, will increase our ability to confidently and richly describe the goals of the activity and application in teaching and learning. Each respondent will provide researchers' with a unique experience and view of the online offering. We have, therefore, sought to address evaluation questions with data from multiple respondents in order to provide data necessary to answer robustly the evaluation questions. For example, student accounts of teaching and learning will help to balance teacher accounts about teaching and learning practices associated with a particular online activity. Exhibit 2 maps the various data collections onto the evaluation questions.

Exhibit 2. Relationship between Data Collections and the Research Questions

Research Question	*Developer Interviews*	*Review of Extant Learning/Use Data provided by the developer or administrators.*	*Product Interface Review*	*Instructor Interview*	*Observation of Use*	*Student Focus Group*
What are the key design features of online learning applications that are congruent with the research base on online learning approaches in K-12 education?	X		X			
What are the key design features of emerging online learning approaches in K-12 education?	X		X			
What user conditions and practices are associated with more successful implementations of online learning applications (replacements for FTF)?*	X	X		X	X	X
What user conditions and practices are associated with more successful implementations of blended online learning applications	X	X		X	X	X

(enhancements to FTF)?*						
What conditions challenge implementation of online learning in the way developers intended?	X			X	X	X

* “Success” is defined primarily on the basis of learning outcomes and secondarily on the basis of usage level (the product of number of users and average hours of use).