



***Conducting and Reporting
Product Evaluation Research:
Guidelines and Considerations for
Educational Technology
Publishers and Developers***

Executive Summary

November, 2011

Author: Denis Newman, Empirical Education Inc.

**A Publication of the
Software & Information Industry Association (SIIA)**

About SIIA

The Software & Information Industry Association (SIIA) is the principal trade association for the software and digital content industries. SIIA provides global services in government relations, business development, corporate education and intellectual property protection to 500 leading software and information companies. SIIA's Education Division serves and represents more than 180 member companies that provide software, digital content and other technologies that address educational needs. The Division shapes and supports the industry by providing leadership, advocacy, business development opportunities and critical market information. SIIA provides a neutral business forum for its members to understand business models, technological advancements, market trends and best practices. With the leadership of the Division Board and collaborative efforts with educators and other stakeholders, the Division undertakes initiatives to enhance the use of educational technology and the success of SIIA members.

Software & Information Industry Association
1090 Vermont Ave, NW, 6th Floor
Washington, DC 20005-4095
Phone: 202-289-7442
Fax: 202-289-7097
<http://www.siia.net>

About the Author and Empirical Education Inc.

Denis Newman founded Empirical Education after a 35-year career in education research and development. His Ph.D. in Developmental Psychology is from the City University of New York. He has conducted research and development at Rockefeller University, University of California, San Diego, Bank Street College of Education and BBN Corporation. He is widely published and has served as program chair for the American Educational Research Association's Curriculum and Learning Division. He was a pioneer in the application of Internet technologies for student learning, professional development and school administration, bringing to market the first integrated web server designed for schools. His business career has included senior positions at educational software companies Tegrity and Soliloquy Learning.

Empirical Education Inc. is an eight-year-old educational research firm in Palo Alto, California, with the mission to help school systems generate evidence to support decisions about products, programs, and policies. The company has a team of 30 research scientists, statisticians, engineers and others with the expertise in and focus on evaluations of instructional, professional development and personnel programs and processes in K-12 schools.

425 Sherman Avenue, Suite 210 • Palo Alto, CA 94306 • (650) 328-1734 • www.empiricaleducation.com

Acknowledgements

SIIA wishes to extend its thanks and appreciation to a number of individuals for their contribution to these Guidelines.

We extend our gratitude to our author, Denis Newman (Empirical Education), for his knowledge, commitment to excellence and persistence in the face of many rounds of discussion and review.

Again, our thanks to Denis and to his co-chair of the SIIA Education Division's Research & Evaluation working group, Rob Foshay (Texas Instruments), for their vision for these Guidelines and their leadership and tireless efforts in guiding them through to completion. Special thanks should also be extended to Mark Schneiderman, SIIA's Senior Director of Education Policy, who convened the group and made the guidelines happen.

We also appreciate these members of the SIIA Education Division's Research & Evaluation working group for their support of this initiative, and for their ongoing input and review:

- Mick Adkisson, SMART Technologies
- Christopher Brown, Pearson Education
- Andrew R. Coulson, MIND Research Institute
- Valerie Crawford, SRI International
- Kristin DeVivo, Scholastic
- David Dockterman, Tom Snyder Productions
- Katie Gilligan, TextHelp
- Carol Guziak, Houghton Mifflin Harcourt
- Dana Jeffrey Laursen, PLATO Learning
- Irene McAfee, McGraw-Hill Education
- Steve Ritter, Carnegie Learning
- Sara Sawtelle, Learning Enhancement Corp.
- Jay Sivin-Kachala, IESD
- Cathy Sump, Curriculum Advantage

Finally, for their thoughtful comments on earlier drafts of this document, SIIA and the author thanks:

- Larry Hedges, Board of Trustees Professor of Statistics and Social Policy, Institute for Policy Research, Northwestern University;
- Robert Slavin, Co-Director of the Center for Research on the Education of Students Placed at Risk, Johns Hopkins University; and
- Talbot Bielefeldt, Senior Research Associate, International Society for Technology in Education.

SIIA and the author take responsibility for all the errors that remain.

Conducting and Reporting Product Evaluation Research: Guidelines and Considerations for Educational Technology Publishers and Developers

Table of Contents

<i>About SIIA – About the Author -- Acknowledgements.....</i>	<i>1</i>
<i>Introduction</i>	<i>4</i>
Purpose and Audience	5
Scope and Limitations.....	5
Challenges Common to Research on Educational Technology	6
<i>Summary and Outline</i>	<i>10</i>
<i>Guidelines and Considerations for Conducting and Reporting Product Evaluation Research</i>	<i>13</i>
Ask the Right Question.....	Error! Bookmark not defined.
Support the Implementation of the Product or Service.....	Error! Bookmark not defined.
Plan a Study of Sufficient Size and Duration to Show an Effect	Error! Bookmark not defined.
Plan for Plausible Causal Claims	Error! Bookmark not defined.
Avoid (the Appearance of) Conflicts of Interest.....	Error! Bookmark not defined.
Provide a Comprehensive and Detailed Research Report	Error! Bookmark not defined.
Make the Research Findings Widely Available.....	Error! Bookmark not defined.
Accurately Translate Research for Customers.....	Error! Bookmark not defined.
<i>Conclusions</i>	<i>36</i>

**NOTE: THIS IS AN EXECUTIVE SUMMARY. IT DOES
NOT INCLUDE THE FULL GUIDELINES (PP 13-35
ABOVE). THE FULL REPORT IS AVAILABLE AT**

<http://bit.ly/vvcVlt>

Introduction

The emphasis on accountability and effectiveness at a time of fiscal constraints has increased the demand by education decision makers for research evidence about their programs, practices, products and services. This includes research on the impact of the educational software, digital content and related technology or technology-based products and services. We can expect this trend to continue among educators, whatever the evolving public policy requirements, because of the increasing availability of data on student and teacher performance on which program evaluations can be based.

In this context the Software and Information Industry Association's (SIIA) Education Division convened a Research & Evaluation Working Group to assist its members in understanding what is being called for and how educators' desire for evidence can be addressed.

Many educational product and service providers have already responded to the need for evidence by enhancing the scale, scope and rigor of their existing research investments, including further documenting the scientific basis of their products and services and commissioning additional evaluation research. While some providers have experienced researchers on staff (some of whom were represented in the SIIA working group that advised on these guidelines), many companies do not have in-house expertise. The working group identified the value of establishing a set of practical guidelines to help in planning for evaluations, to illustrate best practices, and against which to evaluate the potential offerings of outside researchers who might

contract to conduct the studies. The result was this set of 22 guidelines for conducting and reporting evaluation studies of educational technologies. The goal of the working group ultimately was to enhance the quality and credibility and, therefore, the utility of evaluation reports to education decision makers.

The Guidelines are not comprehensive in scale or scope, and they do not constitute a step-by-step manual for designing and conducting evaluation research studies. Instead, they attempt to flesh out four basic ideas that evaluation research must be:

- relevant to educators by providing information needed to inform their decisions;
- transparent and complete in its disclosure and reporting;
- free of undue bias; and
- conducted using accepted research methods.

Although these Guidelines are narrow in scope being geared to K-12 education and technology-based products and services, the principles they embody may apply to other education sectors. The guidelines should not be treated as a handbook for research design or data analysis methods but as a primer of best practices for publisher-developer sponsored or supported studies. The working group also hoped that the document will help inform all stakeholders about the research challenges unique both to studies of technology and to vendor-commissioned research in general.

Purpose and Audience

The Guidelines have several purposes and audiences. First, the Guidelines are intended primarily for publishers and developers (i.e., providers) of educational technologies, including especially software applications, digital content, e-learning and related instructional technology products and services. In particular, the Guidelines are written for the managers responsible for development, evaluation and marketing of these products and services, whether or not they have evaluation researchers on staff. The Guidelines address operational decisions – planning, designing, conducting and reporting – that are under the control of providers carrying out or commissioning a study on their products and services.

To provide information of greatest value, the Guidelines focus on issues that are most unique to educational technology products and services, and are not particularly well addressed in the general literature on evaluation design and methodology. These include determination of appropriate outcome

measures, the importance of fidelity of implementation and issues that are specific to vendor-sponsored research.

In addition, the SIIA working group hoped the Guidelines would also give educators confidence that providers understand the importance of presenting information that is unbiased, actionable and of the greatest value in helping them select and implement technology-based products and services. Research reports that adhere to these Guidelines can be expected to be of high quality and credibility.

Finally, the working group expected that the Guidelines would be reviewed by additional stakeholders – researchers, policy makers and education officials. For all audiences, the Guidelines not only provide a set of standards of practice, but also seek to advance the field by helping to identify the appropriate balance between the rigor, practicality and usefulness of evaluation studies of technology-based products and services.

Scope and Limitations

The Guidelines address only a very narrow type of research: the genre of research that evaluates the impact – that is, the effectiveness or efficacy – of a particular product or service on educational outcomes. This research is inherently a comparison of what happened with a new technology-based intervention to what would have happened if the intervention had not been introduced. Legitimate research design methods distinguish the intervention's impact from the other factors that could have influenced the results, thereby isolating it as the most plausible explanation for the impact. Decision makers may then have confidence that, if they implement the intervention in the manner described in the study, they will

receive an impact similar to that found in the research, given other limitations stated in the report.

Focusing on one kind of research in the Guidelines is not meant to diminish the value of other research genres to education providers and decision makers. Indeed, other research purposes, methods and designs are important to guide product design, selection and implementation. For example, formative research, in particular, has an important role in earlier stages of product development, testing and refinement by asking what it was about the product that made it work, under what conditions it worked and with whom it

worked. These questions may also be addressed by summative evaluation studies, though perhaps at a slightly higher cost. Thus providers would most often move to the kinds of evaluations addressed in these Guidelines only after undertaking a research agenda with other goals and methods.

The Guidelines will not attempt to dictate methodology or to summarize the many volumes written on evaluation research designs. For further information on research designs, readers are encouraged to consult

evaluators with experience in school settings. The Institute of Education Sciences provides a useful registry of evaluators (<http://ies.ed.gov/ncee/wwc/ReviewProcess.aspx>) although they do not certify their relevant expertise. The American Evaluation Association also provides a “Find an Evaluator” page (http://www.eval.org/find_an_evaluator/evaluator_search.asp). Readers may also review SIIA’s *Scientifically Based Research: A Guide for Education Publishers and Developers* (2003; <http://www.siaa.net>).

Challenges Common to Research on Educational Technology

An account of common challenges related to research on educational technology products and services provides an overall frame for the Guidelines. While these issues are not all necessarily unique to this domain, they are challenges typically encountered in conducting such research.

Outcome Measures

Because technology serves many purposes, its impact should be measured in a manner specific to the given type, goals and use of a specific technology. Student achievement and test scores are not the only valid measures. Technology purposes range from instructional to administrative, from assessment to professional development and from data warehousing systems to information productivity applications. The measures could therefore include such outcomes as student test scores, teacher retention rates, changes in classroom practice or efficiency, availability and use of data or other student/teacher/school outcomes that can be observed and measured.

Many of these outcome measures can also be viewed as intermediate outcomes – changes in practice that, as demonstrated by other research, are likely to affect other final outcomes. For example, an evaluation of a certain data system may find its positive impact on the use of data to inform instruction, and we know from other research that this

outcome can help drive improvement in student learning. For purposes of the data system, impact on the use of data is an appropriate evaluation outcome measure, and it should be valued by education decision makers as an intermediary to the ultimate goal of improved test scores.

In addition, it is essential that an evaluation study focused on student achievement select appropriate outcome measures that provide the proper balance between aligning to the specific learning outcomes addressed by the technology and providing generalizability. In some cases, an outcome measure that may be important for school accountability (e.g., state tests) may constitute too blunt an instrument to capture the full value of a certain product or service (e.g., one that is narrowly targeted such that the state test may include curriculum not covered by the intervention). Another assessment may be better aligned to, and therefore better able to measure the impact of, a product or service with a narrower or more

targeted goal – including achievement on a specific set of learning standards, student technology literacy, critical thinking or student motivation. In instances when the comparison group does not cover the material on the specialized test, then we can conclude only

that it possible to teach the material; results of such comparisons can say little about the relative effectiveness of the product studied. In some cases, it will be desirable for the study to use both types of tests.

Fidelity of Implementation

Research results are heavily influenced by the extent and quality of a product's or service's implementation. Educational technology implementation occurs within very complex organizational structure of resources and people. Insufficient hardware access, too little time on task, lack of educator willingness and/or ability to appropriately integrate the technology and inadequate school leadership and support can all negatively affect the implementation and, therefore, the impact. This stands in contrast to some medical trials, where implementation variables depend less on conscious decisions of multiple actors, but are largely based upon whether subjects comply with treatment as prescribed, as well as on biological systems.

Simply providing technology without efforts to measure whether it is being used as intended and is functioning as designed – which may include the vendor ensuring necessary training, support and leadership commitment – may not

be an experimental condition that can be expected to succeed. For example, if a technology is not matched closely to the curriculum and instructional strategy, results are compromised. The condition in the research study is therefore ideally composed not only of the product or service itself, but also of the context and support for its use. In other words, the question is not simply whether the intervention works, but how well it works under particular conditions. Thus the treatment is best described in terms of an implementation model provided by the technology developer. SIIA has provided a useful reference for this: *SIIA Software Implementation Toolkit: Guidelines for Educators* (2007; <http://www.siiia.net>). At the same time, if the amount of support provided for implementation is more extensive than is normally available to customers, the evaluation may become a “hot house” study with more limited generalizability.

Comparison Conditions

In education there is seldom a pure “blinded control” condition such as can be achieved in a medical trial with a placebo or sugar pill, where the placebo is assumed to have no effect, but the subject doesn't know whether or not it is the real medication. In schools, a new math program is typically compared to the math program already in place. Evaluations of education products and services resemble comparative effectiveness trials in medicine in which a new medication is tested against a

currently approved one to determine whether it is significantly better. For any evaluation of a product or service, the measure of effectiveness is really the comparative effectiveness against what is often called the “business-as-usual” condition. Because the effect of the product or service will depend on the existing, or baseline, way of doing things, the same product may prove effective in one district that currently has a weak program but relatively less effective in another where a

strong program is in place. In both cases the technology may have a positive effect, but an impact may not register in the evaluation in cases where it is measured relative to an otherwise effective business-as-usual condition. Thus in education, it is necessary to test products and services in a variety of

settings representing differing comparison conditions. And it is not unreasonable to evaluate a technology in a setting where there is need for improvement; potential customers seeking information on effectiveness are generally those with a problem to solve.

Pace of Research vs. Technology Innovation

Technology products and services are constantly changing and improving. By contrast, in evaluation studies, several years may pass between the initial stage of identifying participants and the final stage of reporting results. In many cases, by the time research is completed, the technology products and services may have been significantly updated and no longer be available in the format or version studied. In this case, educators should appropriately consider studies conducted on previous product versions, as well as those conducted with other populations and in other settings.

Evaluation research, therefore, must be only one of many factors used for decision making. Waiting for comprehensive and definitive research literature on a given intervention will both dramatically limit educators' options and slow the pace of innovation and development. Building rigorous evaluation strategies into earlier field tests and school district pilots are approaches that can expedite product evaluation and help avoid forcing innovation and product development cycles to wait for evaluation research cycles to catch up.

Funding and Review of Product Evaluation Studies

For a variety of reasons, education technology products and services face serious challenges in obtaining both funding and peer review of their evaluation research.

On the funding side, relatively few outside resources are available for product evaluation studies, leaving it to the product developer to fund such studies. Government and foundation resources are limited relative to the large number of technology-based and other educational interventions calling for evaluations. Without support from the publisher or developer, the volume of evaluation research called for by education decision-makers will not get conducted.

On the publication side, research journals do not generally include studies of products

beyond those that are intended to further theoretically-driven research agendas in a given area. Thus the formal research journal is often not the best source of information about technology products and services. Without non-traditional publication channels, the research that is conducted is unlikely to reach the decision-makers.

Education stakeholders, therefore, are likely to find a dearth of independently funded or reviewed product evaluation research. Consequently, a provider must look at the merits of the available research and, while following the guidelines presented here and not making claims that are unsubstantiated by the evidence, highlight studies that will be useful for decision-makers in forming an

initial impression, even if they consist only of case studies of successful implementation.

Following these guidelines, the provider can view any sale as an opportunity for collecting systematic data. For example, building on pilots paid for by the schools, the cost of research can be reduced. Often federal or state grants call for an evaluation, although often the RFP's evaluation specifications (or agency enforcement of the stated criteria) do not call

for the level of rigor suggested in these guidelines as needed for conclusions of effectiveness. In this case, the publisher will have to persuade the customer to exceed the RFP requirements so as to produce evidence of the product's effectiveness in the district. This example illustrates a major challenge to providers in conducting and reporting research, which is the context for these guidelines.

Summary and Outline

The following specific guidelines comprise the core purpose of this document – to describe standards of best practice for the conduct and reporting of evaluation research on technology-based products and services. Below are the guidelines in outline format. Following the outline, the remainder of the document will fully describe these guidelines and related considerations.

Ask the Right Question

1. Match the research question (and, ultimately, its outcome measure) to the intervention's goals, including where appropriate, intermediary goals such as a change in practice that the research literature suggests enhances achievement or other important outcomes.
2. Select outcome measures that provide an appropriate balance between being sufficiently sensitive to the particular outcomes targeted by the product or service (e.g., a subset of learning standards), and aligning to a more general measure used for educational accountability (e.g., high-stakes state tests).
3. Before the study begins, decide whether to evaluate the product or service in one of two ways: (a) under ideal, "hothouse" conditions (i.e., an efficacy study); or (b) under ordinary field conditions, where an impact may be more difficult to discern (i.e., an effectiveness study).

Support the Implementation of the Product or Service

4. Develop and document as explicit a model as possible for how to implement the product or service in the educational setting. This includes the appropriate technology infrastructure, educator training and product usage required for an impact to be detected. The more explicit this model, the more likely the research will be able to explain the results in terms of whether the implementation met these expectations.
5. In conducting an evaluation, distinguish between correlational and causal findings. It can be useful to check for correlations in the data, such as between implementation fidelity and outcomes, for the purposes of product improvement and understanding best practices. But be cautious in drawing conclusions about a causal effect of the intervention from correlational findings, as a factor other than the intervention can often provide a plausible explanation.

Plan a Study of Sufficient Size and Duration to Show an Effect

6. Establish the study's "unit of analysis." This is the sample unit level – typically school, teacher or student – at which the product or service is designed to be used. The appropriate unit may be determined by the implementation model, as when the model requires treatment to be administered school-wide. Otherwise, the unit of analysis may be determined based on cost constraints. For example,

it costs less to randomize 40 students than 40 schools to treatment and control conditions.

7. Employ a sample size sufficiently large to draw conclusions with statistical confidence, taking into account the magnitude of the expected effect, the availability of a pretest and the number of units of analysis needed.
8. Plan for a study of sufficient duration for the product or service to have its intended effect. Consider the period needed for training and other start-up activities, and allow time for full integration into instructional and administrative processes.
9. Identify the comparison condition or clearly defined baseline relative to which the estimated impact of the evaluated product or service is measured. The comparison condition is needed to determine what would have happened without the new product or service.

Plan for Plausible Causal Claims

10. Choose a research design that is capable of reducing plausible alternative explanations for changes in performance, other than the impact of the product or service under study.
11. Avoid or mitigate selection bias in identifying the group that uses the new product or service and the group to which it is compared. A method to be considered is random assignment of study units (e.g., school, teacher or student) to use the intervention. Where random assignment is not feasible, other approaches to identifying a well matched comparison group can be used.

Avoid (the Appearance of) Conflicts of Interest

12. Follow standards of practice and regulations put in place to protect the privacy and safety interests of study participants. These often include review by an Institutional Review Board and adherence to the Family Educational Rights and Privacy Act (FERPA).
13. Work with researchers who can be objective and independent. Take steps in selecting the researcher, determining the editorial and reporting process and funding the study that will help ensure objective findings. This applies whether the research is conducted internally or through an external contractor.
14. Design participant incentives to avoid any bias in the results. While teachers and other participants are commonly offered honoraria and other benefits, excessive inducements, especially if they favor the group using the product or service, may influence the results and should be avoided.

Provide a Comprehensive and Detailed Research Report

15. Produce a full research report that thoroughly describes the research conditions and context in detail, including the product or service, its implementation, group

assignments, comparison conditions, populations, interactions and any factors that may cause bias. Only a sufficiently detailed report allows for a third party to evaluate its conclusions and, potentially, to replicate the study.

16. Distinguish between (a) the findings pertaining to the original core hypothesis and (b) the exploratory results and conjectures arising from post-collection review of the data.
17. Be clear about the study origins, initiating parties and funding sources.
18. Be clear about study authorship and final editorial control.

Make the Research Findings Widely Available

19. Make the research report available through a variety of channels, such as a refereed (peer reviewed) journal, conference presentations, research clearinghouses and the company website.
20. Make all formal evaluation research findings available upon request regardless of the outcome, except in these instances: (a) a “failed experiment” where it is determined prior to review of outcomes data, for example, that the product or service was not implemented with fidelity, too few participants could be recruited, the study was too poorly designed or the data could not be collected; or (b) determination by the provider that the product or service must be improved and re-released, in which case the results can be considered as formative information for product improvement.

Accurately Translate Research for Customers

21. In the marketing literature for a product or service, accurately describe its impact – relative to the strength of the research design, quality of the evidence and size of the effect – using language that educators without research training can understand.
22. Cite the full research report any time the research or its findings is referenced.

Conclusions

Educational technology providers engaging in research is not new. What is new is the relative attention now being paid by education decision makers to the research basis of product effectiveness. These Guidelines attempt to outline several practical considerations and best practices most unique to technology and to provider-sponsored research. Because the Guidelines are limited in scale and scope, they do not constitute a how-to manual, address all issues or cover issues in-depth. Instead, they are intended to be used to help providers ensure the quality of the evidence on their interventions available to customers.

The hope is that the Guidelines will help providers in understanding research as an ongoing process, rather than a one-time activity. This is especially important in light of the speed of technology innovation and new product development, which will often outpace the research cycle and educators' calls for evidence of effectiveness. Traditional tools for review and dissemination of research, while still an important avenue, are often not able to keep up with new versions of products or services.

In addition, because the educational technology industry is faced with a very diverse marketplace, effectiveness research for their products and services cannot be conducted for every local curriculum and district population type. A reasonable sampling of contexts should be provided as educators look for a close match to their own criteria, all the while recognizing that an exact match may not exist. Alternatively, cooperative research between providers and customers can afford the local evidence that educators need to support their decisions. Such public-private partnerships are critical to meet education's needs. In fact, if the marketplace has to wait for providers to make localized evidence available before products and services can be implemented, innovation will be stalled as development cycles are forced to slow down to wait for evaluation research cycles to catch up.

The Guidelines provided here do not offer all the answers to how research can be financed or how it can provide timely answers for educators facing serious educational challenges and seeking effective solutions. The goal is to help providers understand the basic standards of research practice required for evidence that educators can use and to find productive and workable approaches to conducting and reporting that research.