# The enthusiasm effect, the reality effect and other things to know when evaluating professional development impacts

Jason Ravitz <jason@bie.org>
Buck Institute for Education
18 Commercial Blvd., Novato, CA 94949
United States
415-883-0122 x 310
www.bie.org

**Keywords:** Evaluation methods, longitudinal data, teacher change, technology, constructivism

## Abstract

A statewide evaluation in Idaho addressed changes in teaching philosophy and technology skills that occurred during summer in-service workshops. This paper discusses how to think about changes in teachers' technology skills and teaching beliefs and whether these sustained themselves in the face of school year and classroom realities. . It also shows the importance of disaggregating data to see for whom and under what conditions the changes were most substantial. The results provide guidance for evaluating the impacts of professional development. Teachers' beliefs about the efficacy of teaching reforms and their own technology skills varied at three *different times*: prior to summer workshops, after the summer workshops, and after ten months of subsequent teaching. Initial changes reported at the conclusion of a workshop represent an “enthusiasm effect” and changes after teachers have tried new ideas and new technologies in the classroom represent a “reality effect."

**Overview**

This paper provides a number of ideas that are important for professional developers and evaluators to consider. This section provides an overview of steps that can be taken to improve confidence in findings. Examples are provided from several evaluation studies in Idaho (Buck Institute for Education, 2002a; 2002b).

First, it can be helpful to have a theory of change for evaluating professional development, especially in the complex field of educational technology (Haertel & Means, 2000). It is often a good first step to understand program goals clearly, and to plan accordingly. As discussed below, program theory ended up being important for interpreting findings in Idaho.

Second, this paper shows the importance of creating a good baseline measure prior to program participation; it also provides some key characteristics such a measure. It is important to control for initial scores before trying to evaluate change. One approach is to use standardized residual scores (provided as an output of multiple regression in statistical packages like SPSS). These provide a useful measure of changes in scores between two times and can easily be visualized.

A third critical step is to identify systematic sources of error, or bias, that can influence findings. It is extremely informative to focus on *for whom and under what conditions* professional development yields the greatest benefit. The same can be said for technology impacts on learners. Descriptions of one variable at a time and bottom-line comparisons are rarely useful without some comparison to other people or times.

Fourth, it is important to recognize that teacher change in educational technology is not a one dimensional construct. It is probably much easier to develop technology competencies than it is to change beliefs about teaching and learning. Change scores on the constructivist belief measures are very small compared to the gains in technology skills that are reported. Given the emphasis of TWT on reforming teaching practices, these results are important. They are also consistent with earlier findings showing the limited relationship of technology skills on their own to influence classroom practice (Ravitz, 1999)

Fifth, it is important to consider the timing of data collection, as suggested in the paper title. A baseline measure prior to participation can help account for individual and group differences that were present at the outset. Observations or measures after the workshop can show if professional development played a part in changing teachers' outlook, skills or practice. This can be shown right away. If ideas have not taken hold or changed during the workshop, it is doubtful they will afterwards.

Finally, it is important to recognize that constructivist-related reforms and technology applications are often more difficult to implement in the classroom than they seem in a workshop. In a number of cases enthused teachers may report a belief that is consistent with constructivist learning theories presented in a workshop, but find this belief less tenable or more difficult to put into practice in their actual classrooms. They also may report they have mastered a new technology and later find it necessary to receive much more support. Pre-post research designs that do not leave time for "reality" to set in may obtain artificially favorable results due to what we are calling an "enthusiasm effect." It helps to have a follow-up measure or observation to see how enthusiasm generated or displayed at a workshop holds up in the face of reality. If the initial gains were substantial there are still likely to be overall gains,, even if there was some drop in the self-ratings of teachers after the workshop This was the cases for most of the items in the Teaching with Technology (TWT) study described in the next section.

**Project summary**

Teaching with Technology (TWT) is a statewide professional development program in Idaho funded and developed by the J.A. and Kathryn Albertson Foundation (www.jkaf.org). TWT provides intensive summer workshops led by teachers who provide follow-up support as Technology Fellows during the school year. Each workshop lasts five days and covers basic technology skills, advanced technology applications and several constructivist-oriented reform models for teaching. It was a natural complement to the Opportunity One program that had provided funding for technology equipment for use across Idaho's schools in prior years (Ravitz, Mergendoller & Rush; 2001; BIE, 2002a).

For our evaluation of TWT as a professional development program we used multiple indicators of program impact including reported helpfulness of training workshops and *changes* in teachers' attitudes toward technology; requests for technology training; objectives for student technology use; beliefs about teaching and learning; and in teachers' technology skills (BIE, 2002b).

**Study methods**

This paper identifies and treats several sources of systematic error, or bias, that can influence the evaluation of professional development outcomes. We focus on initial scores as a predictor of subsequent scores, the timing of data collection, and differences by grade level, and school size. By including these variables in our analyses, we improve confidence in our understanding of professional development impacts.

We used surveys administered at three different times to all participating teachers. Prior to holding the professional development workshops, TWT sent self-report surveys to all registered participants via mail; if teachers did not return the surveys they were asked to complete the instrument when they entered the workshop. A second wave of surveys was given to teachers as they left the workshop to rate their initial responses and changes in their points of view and perceived technology skills after five days of intense professional development. A third wave of surveys was administered via an online survey submission form 10 months later, i.e., after a year of teaching.

There were differences in the survey administration and scope, but key variables concerning teaching beliefs and technology skills were measured the same way during each administration. The survey included measures of teaching beliefs that broadly indicate a constructivist-oriented view of teaching and learning. These items correlate with classroom practices and have strong reliability in our study and previous studies, standardized reliability alpha > .80 (Ravitz, Becker & Wong, 2000, p. 4-5; BIE, 2002a, BIE 2002b). The technology skills measures were even more reliable but in the past these have tended to predict technology use by teachers on their own professionally more than with students (Becker, Ravitz, & Wong, 1999; Ravitz, 1999).

Although the number of respondents in Wave 3 was relatively low compared to Waves 1 and 2, the differences in their responses at this time were substantial and provide the basis for this report. We used effect size differences to rate the amount of change in scores among participants between waves, using the Wave 1 standard deviations to compute the effect size of the change (Thompson, 1999). In order to maximize the power of our longitudinal data we used paired-t tests to determine changes over time. The trade-off of this approach is that it limited the number of cases that were available to analyze. With more cases available, we would have liked to include within and across subject comparisons (BIE, 2002a).

We acknowledge that TWT teachers were self-selected and started with more technology skills and with slightly more constructivist beliefs than the average Idaho teacher (BIE, 2002b). We cannot say how non-participating teachers would have fared had they chosen to participate in TWT, or how non-respondents might have responded to our surveys. As is always the case, we ultimately have to rely on replications to provide true confidence in our findings (Katzer, 198, Krathwohl, 1998).

**We must theorize about the relationship of technology and pedagogical change**

This section presents different ways of thinking about how technology skills and constructivist beliefs are related and how they are developed among teachers. How we view these relationships and our "theory of change" will impact how we evaluate our work. Professional developers and their evaluation teams should have a clear idea about who is supposed to change, how they are supposed to change, and under what conditions.

Often teachers who report more technology skills also tend to report more constructivist-compatible beliefs about teaching and learning. This is not just because different subjects and levels vary in their patterns of teaching and technology use (Ravitz, 2002). Even compared to colleagues who teach the same grade and subjects, high-end technology using teachers tend to be more constructivist in their beliefs than others (Ravitz & Mergendoller, 2003).

New technologies may just be more attractive to constructivist-oriented teachers who see the advantages of recent applications for engaging students in meaningful and challenging activities (Becker, 2001; Jonassen, 2000, GLEF, 2003) advantage of various software more than others, *without it fundamentally changing their beliefs about teaching.* In this case, the development of technology skills might have the greatest impact on already-constructivist teachers. There may also be a third variable – e.g., their academic training emphasized both, their professional orientation and school culture encourages both (Becker & Riel, 2000) or just plain innovativeness (Rogers, 1983) or "dissatisfaction with the status quo" (Ely, 1990) - that causes these teachers to seek new practices and technologies that will improve their practice.

A more hopeful possibility for reformers is that professional development can actually stimulate new ideas in teachers by making them take seriously a reform-oriented view of learning and allowing them to see how it might work in their classrooms. When they see new ways of teaching and learning and what students can do with technology they may decide that it is preferable (and possible) to adopt new pedagogy with the help of technology.

The suggestion that technology can change teachers has some support. Becker and Ravitz (1999) found that teachers who were early Internet users reported changes in their practices and attributed them to the technology. In our view, this is the type of impact that programs like TWT want to have; they want to see technology being used as a catalyst for teaching reforms. Our evaluation was guided by this understanding that TWT sought to change long-term beliefs about teaching and to develop teachers' efficacy with technology, with the emphasis on the former. Importantly, the teaching changes and uses of technology that are being described have been linked to learning gains for students (BIE, 2002a; George Lucas Educational Foundation, 2003; Mergendoller, Maxwell & Bellisimo, 2000). This topic is discussed further in the next section.

**Where *students* end up is largely (but not entirely) a function of where they started**

It is unheard of for large groups of learners to all enter a class, lesson or workshop with the same background knowledge and experience. It is essential to get a sense of how teachers or students differ at the start of an intervention or workshop before assessing their individual progress. Starting with strong test scores and ending with strong test scores may not represent much learning at all. Unless there is contrary evidence, the higher scores on the pre- and post-test are probably a characteristic of the student before the year began. The same concern appears later with respect to the self-report data from teachers.

Controlling for initial scores is important because we often anticipate that student achievement is related to background variables in the students' lives (e.g., parental education and income); these are un-seen and un-measured background variables that are typically not available for analysis. Controlling for prior achievement goes a long way towards creating “all things equal” conditions, without randomization. *The original score takes into account any sources of systematic error or bias* besides what happens during the intervention and looks at how students progressed from there. This, in effect, controls for all the unseen variables that may have accounted for the initial score. Laczko-Kerr & Berliner (2002) and Russell (2000a; 2000b) discuss alternate approaches to this problem.

Figure 1 shows that prior performance is generally the best predictor of future performance. It shows Language Arts scores of 8$^{th}$ graders in Idaho on the Iowa Test of Basic Skills (ITBS) for two consecutive years. The 1999 test score accounts for nearly 80% of the variance (r-squared) on the 2000 test; this is conservative because we used a restricted range. Using extreme scores for the grade, the amount of variance accounted for was closer to 90%.

**Figure 1: Scatterplot between 1999 and 2000 test scores, ITBS Language Arts**

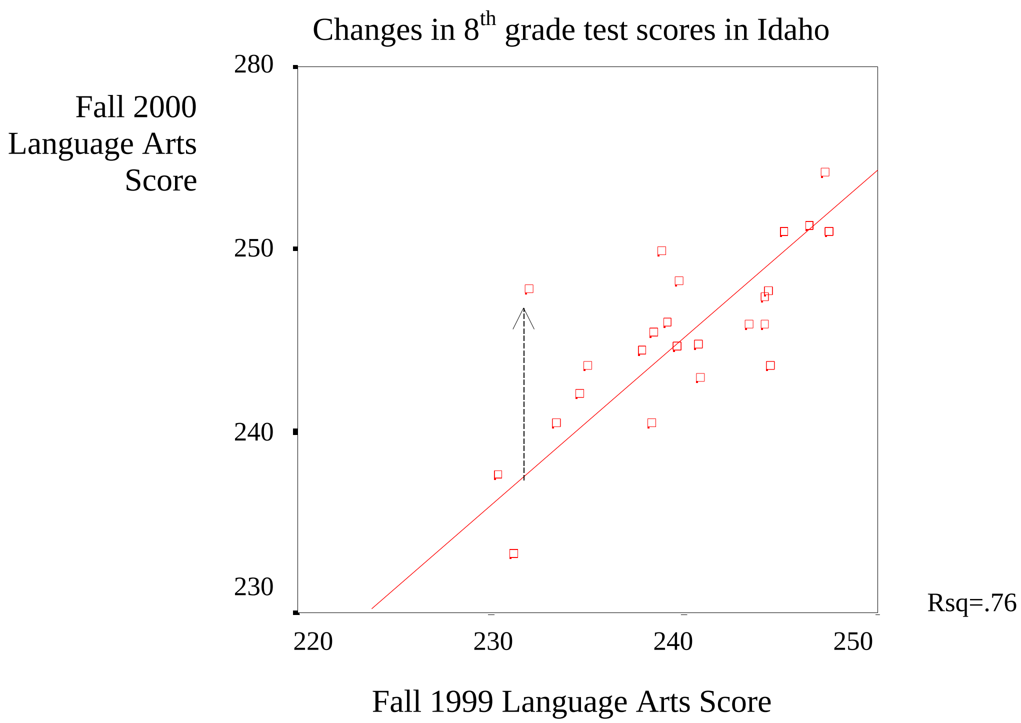



Figure 1 shows how students scored in 2000 (on the y-axis), and how they scored in 1999 (on the x-axis). The prediction line (the solid diagonal line) shows the line of “best fit”, i.e., what would be predicted for each student based on their initial scores and how everyone else scored. A point that falls directly on the prediction line represents a student who gained the average amount, about five points in this case. Points above the line represent gains more than expected; points below the line represent gains less than expected.

The vertical distance above or below the prediction line (broken line) is the “residual" after taking into account the original score and the average amount of change. These scores are particularly useful when everyone’s score is expected to change somewhat. Comparisons have to be made concerning the amount of change, not just whether there was change. This is the case with tests given over several years (Figure 1) and is true on the self-report measures from teachers, shown below.

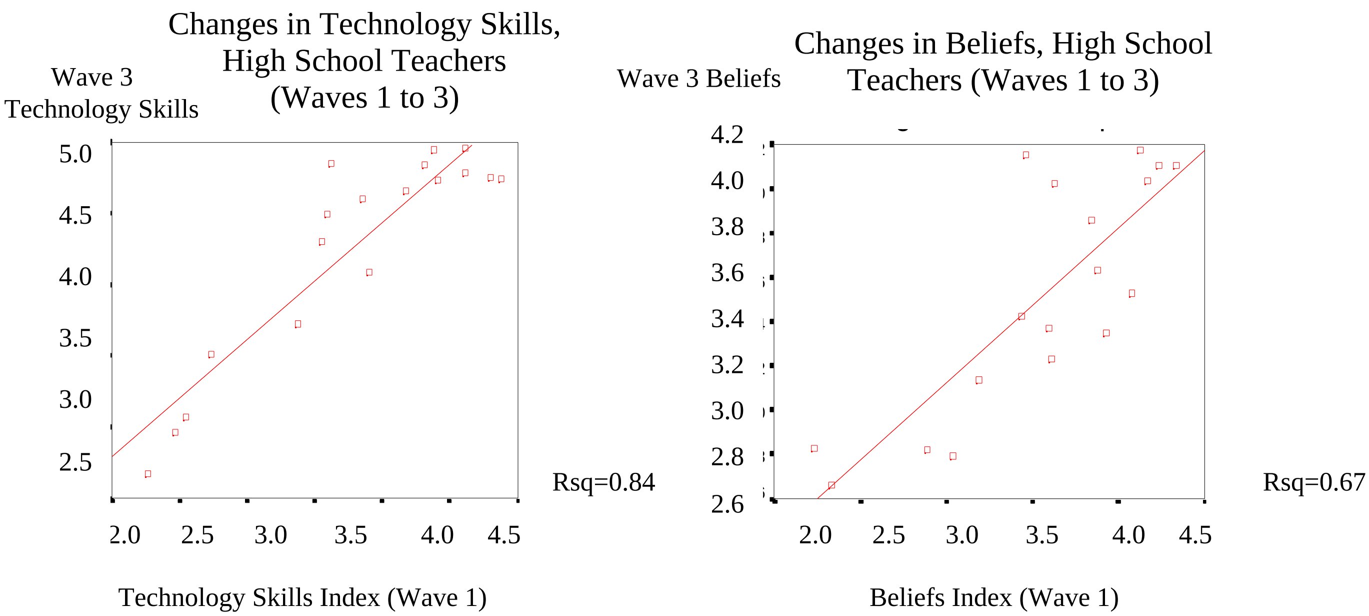
In the highlighted case in Figure 1, one student had the third lowest score in 1999 (starting just above 230) and a year later this student one of the highest scores (almost 250). Others who started near 230 made no such gains. The gain for this student is far more impressive than the gain of the students who ended up with scores closer to 260. The increase in scores for this one student could be a chance fluctuation, but if there are consistent findings about who gained more than others, this could be important, especially if the same pattern was observed across numerous settings. In BIE (2002a) we used residual gain test scores on the Iowa Test of Basic Skills (ITBS) to compare students, teachers and schools. Technology skills were associated with not only higher test scores, but greater gains as well (see also Ravitz, Mergendoller & Rush, 2002).

**Where *teachers* end up is largely (but not entirely) a function of where they started too**

The rest of this paper addresses changes in teachers' self-reported technology skills and teaching beliefs. Importantly, we used reliable measures that were evenly distributed and that left room for increases and decreases in scores. It is common for teachers to hit the “ceiling” on a single survey item by giving the most extreme response that was offered, but index scores can provide room for change and spread scores out very effectively (Ravitz,1999).

Figure 2 reveals that teachers in TWT who started with low technology skills ended with low technology skills; the same is true for constructivist-compatible beliefs about teaching; but once again the correlation is not perfect. Just like the student example above, points above the line represent teachers whose self-report scores gained more than others from where they started. Data about which teachers gained more is of considerable value, and is shown later in Table 1.

**Figure 2. Scatterplot of Teacher Pre- and Post- Ratings of their Skills and Beliefs**




Also of interest is whether those who gained more than expected on technology skills also reported more changes in their beliefs. We computed the correlation between the residual gains scores for technology skills and the residual gain scores for constructivist beliefs. For elementary teachers there was no relationship between who gained on each measure (r = -.17, NS, n=63), for high school teachers there was a correlation, but not a statistically significant one (r = .25, NS, n=18). For middle school teachers, however, there was a statistically significant correlation in residual gain scores (r = .48, p < .01, n=30).

Our tentative interpretation is that to the largest extent the program worked as intended for middle school teachers. It was in this grade level where we simultaneously saw change in both technology skills and in beliefs about teaching and learning. In the other grade levels, teachers were as likely to gain on one measure and not the other, or to report no change at all.

**Change is not uni-dimensional**

It seems to be much harder to change beliefs about good teaching and learning than to provide technology competencies. Despite the best intentions of TWT organizers, many participants developed technology skills to a large extent without changing their beliefs about teaching.

Table 1 provides an overall comparison between responses during Wave 1 (prior to training) and Wave 3 (after 10 months of teaching). Across the board there were substantial increases in the technology skills reported by teachers and only moderate increases in their constructivist beliefs, if any. Changes in beliefs about teaching and learning were in a constructivist direction, but they were relatively small. The average effect size difference in constructivist beliefs was less than 1/5 a standard deviation (ES=.17) while the effect size difference in reported technology skills was more than a standard deviation (ES=1.51).

Table 1 demonstrates the limitations of simple descriptive reporting of data, that is, one variable at a time without reference to others. A description of Wave 3 findings by themselves, without reference to Wave 1 would show that teachers in the larger high schools (bottom row) had the strongest technology skills and the most constructivist beliefs at the end of the study. Without looking at Wave 1, you would not realize that was a characteristic of the teachers in large high schools before the workshop, and is in no way an indication of program impact. Looking at the change scores, high school teachers actually changed the least in their belief about teaching, if at all. Only teachers in elementary schools and in smaller middle schools reported sustained changes in beliefs through Wave 3. This cannot be observed from Wave 1 or Wave 3 scores alone.

Finally, Table 1 illustrates the value of disaggregating teachers' scores so that self-reported data is presented according to the key characteristics of the teachers and schools that may influence these scores. Katzer (1981) explains that it is preferable to treat important and known sources of bias, as variables of study, as opposed to converting these sources of bias to noise through randomization.

**Table 1. Changes in Beliefs and Skills, paired t-tests by school size and grade**

| | | **Beliefs Index (higher score = more constructivist beliefs)** | | | **Technology Skills Index (higher score = greater technology skills)** | | |
|---|---|---|---|---|---|---|---|
| **Teachers in Sample** | **N** | **Wave 1** | **Wave 3** | **Effect Size** | **Wave 1** | **Wave 3** | **Effect Size** |
| **ALL Teachers** | 119 | 3.45 | 3.55 | .17*** | 3.16 | 4.19 | 1.51** |
| | | .58 | .60 | | .68 | .72 | |
| **<u>By Grades Taught</u>** | | | | | | | |
| Elementary school | 63 | **3.42** | **3.60** | 0.36*** | **3.05** | **4.09** | 1.66** |
| *S.d.* | | *0.51* | *0.65* | | *0.63* | *0.73* | |
| Middle school | 30 | **3.33** | **3.44** | 0.17 | **3.22** | **4.27** | 1.49** |
| *S.d.* | | *0.66* | *0.62* | | *0.70* | *0.66* | |
| High school | 18 | **3.56** | **3.51** | -0.07 | **3.28** | **4.26** | 1.18** |
| *S.d.* | | *0.62* | *0.52* | | *0.83* | *0.74* | |
| **<u>By Grades and Size</u>** | | | | | | | |
| Smaller middle schools | 9 | **3.41** | **3.67** | 0.40* | **2.89** | **4.00** | 1.42** |
| *S.d.* | | *0.64* | *0.52* | | *0.78* | *0.79* | |
| Larger middle schools | 19 | **3.44** | **3.44** | 0.00 | **3.41** | **4.42** | 1.60** |
| *S.d.* | | *0.60* | *0.60* | | *0.64* | *0.56* | |
| Smaller high schools | 7 | **3.56** | **3.43** | -0.17 | **2.96** | **3.83** | 0.93** |
| *S.d.* | | *0.73* | *0.60* | | *0.93* | *0.73* | |
| Larger high schools | 7 | **3.75** | **3.72** | -0.12 | **3.66** | **4.75** | 2.21** |
| *S.d.* | | *0.29* | *0.35* | | *0.49* | *0.18* | |

*** p < .001, ** p < .05, * p < .10

<u>Note.</u> Effect sizes are based on Wave 1 standard deviations. Item scores ranged from 1 to 5.

If you randomly took the responses of teachers and did not control for grade and school size, the effect size you would get is the overall effect, show in the first row of Table 3. You would not have the benefit of seeing under which conditions there was change. This is why calls for randomization are often tempered by calls for carefully matched and control group designs. Randomization is best used to rule out sources of bias that you *did not or could not afford to treat* (Katzer, 1981). If you know grade and school size is important, you should include these variables in your study. The same is true for other background variables that you believe might bias results.

**It is helpful to look at initial changes (enthusiasm) and changes after some time (reality)**

In this section we look at the timing of data collection as a final source of error in interpreting professional development outcomes. We noticed that on average the changes that were reported immediately after the workshop (Wave 1 and 2 changes) seem to reflect an "enthusiasm effect" causing them to be somewhat inflated compared to what we saw at the end of 10 months of teaching. We consider this movement back towards the original scores to be a “reality effect” that is probably based on difficulties encountered after the workshop and in the face of real teaching. This does not mean there was no real change during or as a result of the five day workshop. In the case of truly difficult and complex changes, we could say that after a large gain in self-reported skills or beliefs some adjustment back to “reality” is to be expected, as noted by Rockwell & Kohn (1989).

Changes in beliefs across all 3 waves are shown in Table 2. The second column in Table 2 shows the type of teaching belief item and whether the proposed teaching beliefs were conceptualized as more or less constructivist-oriented (C) or traditional (T). For example, the first row shows that teachers initially moved away from the “traditional” view about the importance of structured instruction (ES=-0.28). The negative sign means they moved away from the belief. From Wave 2 to Wave 3 there was only a minute change back towards this traditional belief about the efficacy of structured instruction (ES = 0.04). The result is that the initial change that seems to "hold up" and the overall change was substantial between Wave 1 and 3 (ES=-0.47). In contrast, for the responses concerning the importance of having a quiet classroom and valuing of student initiative, we see a stronger "reality effect." The last column in Table 2 shows what the overall change was for each item between Wave 1 and Wave 3. These were all in the intended direction, if only slightly.

**Table 2. Effect size changes in teaching beliefs between waves, with paired t-tests**

| Statement of Belief | Type of Item | Waves 1 – 2 Enthusiasm effect after training (N=196) | Waves 2 – 3 Reality effect after 10 months (N=84) | Waves 1 - 3 Overall change for entire study (N=114) |
|---|---|---|---|---|
| Structured instruction is important | T | Away<br>-.28*** | *Slightly towards*<br>*0.04* | Away<br>-0.47*** |
| Simple problems to keep class attention is best | T | Away<br>-.25*** | Slightly away<br>-0.10 | Away<br>-0.30*** |
| Students should establish criteria for assessments | C | Towards<br>.34*** | Slightly towards<br>0.12 | Towards<br>0.31** |
| Mr. Jones' more constructivist approach is more comfortable | C | Towards<br>.44*** | *Slightly Away*<br>*0.09* | Towards<br>0.26** |
| Important to keep whole class on single task | T | Away<br>-.20** | *Slightly towards*<br>*0.07* | Away<br>-0.16* |
| Important to have a quiet classroom | T | Away<br>-.30*** | *Towards*<br>*0.30*** | Away<br>-0.15 |
| Students should share work outside of class | C | Slightly towards<br>0.12 | Towards<br>0.18 | Towards<br>0.15 |
| Valuing of student initiative in the classroom | C | Towards<br>.36*** | *Away*<br>*-0.15* | Towards<br>0.12 |
| Teacher should make all the decisions | T | Slightly away<br>-0.12 | *Slightly towards*<br>*0.02* | Away<br>-0.12 |
| Mr. Jones' more constructivist approach produces better knowledge | C | Towards<br>.31*** | *Away*<br>*-0.12* | Towards<br>0.08 |

*** $p < .001$, ** $p < .01$, * $p < .10$   *Italics* show reversal of direction. Effect size shows the magnitude of the change using Wave 1 standard deviations.

After ten months of teaching, teachers valued a quiet classroom more and rated student initiated lessons as less valuable than just after the workshop. Nevertheless, the overall change from Wave 1 to Wave 3 was in the intended direction. For another item, shown in the bottom row of Table 2, the initial gains (or shifts towards the more constructivist-compatible beliefs) were almost entirely erased by Wave 3 --- regarding beliefs about knowledge produced in a classroom taught by a more constructivist-oriented teacher, Mr. Jones (Ravitz, Becker & Wong, 2000).

Table 3 demonstrates that the same "enthusiasm" and "reality" effect can be seen, but are seen less frequently, in teachers' self-ratings of their technology skills. A negative sign in Table 3 means fewer skills were reported than had been reported previously.

**Table 3. Effect size changes in technology skills between waves, with paired t-tests**

| **Self-reported technology skills to…** | **Waves 1 – 2 Enthusiasm effect after training (N=142)** | **Waves 2 – 3 Reality effect after 10 months (N=57)** | **Waves 1 - 3 Overall change, for entire study (N=114)** |
|---|---|---|---|
| Develop multimedia presentation | .81*** | *-0.13* | .69*** |
| Create web page | .78*** | -*.31*** | .55*** |
| Prepare slide show using presentation software | .52*** | 0.09 | .49*** |
| Attach files to email | .28*** | .24** | .47*** |
| Troubleshoot network problems | .54*** | 0.25 | .46*** |
| Display Directory of Disk | .34*** | 0.06 | .37*** |
| Copy Files between Disks | .30*** | .16* | .36*** |
| Create wp document with graphics | .31*** | .18** | .32*** |
| Create Database | .22*** | *-0.11* | .28*** |
| Use WWW search engine | .12* | 0.11 | 0.12 |
| Create Spreadsheet to calculate grades | .16** | 0.11 | 0.10 |

*** $p < .001$, ** $p < .02$, * $p < .10$

We see this "reality effect" drop in reported technology skills between Wave 2 and Wave 3 for just a few types of software. Teachers were less confident in their skills to create a web page and in their multimedia skills than they were at the end of the workshop. Despite the fact that these ended up being the same skills with the greatest gains reported overall (last column), this drop from Wave 2 to Wave 3 probably represent a "correction" that reflects the complexity of using these technologies while teaching.   Other skills continued to increase over time (e.g., attach files to email), suggesting these skills are perhaps more easily obtained and put to use.

Simply put, some technology skills probably can be learned in a few hours or days, while others may take years to develop..  This does not mean these skills should not be taught, but it is important for professional developers and evaluators to be aware of skills where teachers may require ongoing support.

**Summary**

This paper has offered insights from evaluations conducted in Idaho that can inform professional developers, researchers, and evaluators.  Evidence has been provided concerning the usefulness of: 1) judging impacts based on a theory of what was supposed to happen; 2) establishing baseline measures to see where people started; 3) controlling for where people start before judging gains; 4) identifying and disaggregating data to account for systematic sources of error that will bias your findings, like school grade and size; and, 5) seeing how initial reactions to workshops change over time.

The TWT evaluation was conducted under unique circumstances in Idaho, but the methods presented here can be useful and the findings could be replicated in many settings. Hopefully this paper will contribute to ongoing discussions about how to measure professional development impacts in educational technology programs.

Author's note:  The author wishes to acknowledge the contributions of John Mergendoller and Claire Adams of the Buck Institute for Education in the preparation of this paper.

**REFERENCES**


Becker, H. (2001). How are teachers using computers in instruction. Paper presented at the 2001 Meetings of the American Educational Research Association. [WWW Document]. Available: http://www.crito.uci.edu/tlc/FINDINGS/special3/

Becker, H. & Ravitz, J. (1999). The Influence of Computer and Internet Use on Teachers' Pedagogical Practices and Perceptions. *Journal of Research on Computing in Education, 31*(4).

Becker, H., Ravitz, J. & Wong, Y. (1999). *Teacher and teacher-directed student use of computers and software. Report #3. Teaching Learning and Computing: 1998 National Survey.* Irvine, CA: University of California at Irvine, Center for Research in Information Technology in Organizations. [WWW Document]. Available: http://www.crito.uci.edu/tlc/findings/ computeruse/

Becker, H. & Riel, M. (2000). *Teacher Professional Engagement and Constructivist-Compatible Computer Use Report # 7. Teaching Learning and Computing: 1998 National Survey*. Irvine, CA: University of California at Irvine, Center for Research in Information Technology in Organizations. [WWW Document]. Available: http://www.crito.uci.edu/tlc findings/report_7/startpage.html

Buck Institute for Education (BIE, 2002a). *Opportunity One: Technology initiative evaluation for the J.A. and Kathryn Albertson Foundation* [WWW Document]. Available: http://www.bie.org/research/tech/large-albertson.php

Buck Institute for Education (BIE, 2002b). *Teaching with Technology: A statewide professional development program evaluation for the J.A. and Kathryn Albertson Foundation.* [WWW Document]. Available: http://www.bie.org/research/tech/twtfinal.php

Ely, D. (1990). Conditions that facilitate the implementation of educational technology innovations. *Journal of Research on Computing in Education, 23*(2), 298-305.

George Lucas Educational Foundation (GLEF, 2003). Problem based research summary. [WWW Document]. Available: http://glef.org/PBL/research.html

Haertel, G., & Means, B. (2000). *Stronger designs for research on educational uses of technology: Conclusion and implications.* Menlo Park, CA: SRI International. [WWW Document]. Available: http://www.sri.com/policy/designkt/found.html

Jonassen, D. (2000). Computers as Mindtools for Schools: Engaged Critical Thinking. Upper Saddle River, NJ: Prentice-Hall, Inc.

Katzer, J. (1981) Understanding the research process: An analysis of error. In: Busha, Charles H., Ed. *A library science research reader and bibliographic guide*. Littleton, CO: Libraries Unlimited, pp. 51-71.

Krathwohl, D. (1998). *Methods of Educational and Social Science Research (2nd Ed.)*. New York, NY: Addison- Wesley.

Laczko-Kerr, I., & Berliner, D.C.. (2002, September 6). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy," *Education Policy Analysis Archives, 10*(37). [WWW Document]. Available: http://epaa.asu.edu/epaa/v10n37

Mergendoller, J. R., Maxwell, N. L., & Bellisimo, Y. (2000). Comparing problem-based learning and traditional instruction in high school economics. *Journal of Educational Research, 93*(6), 374-382.

Ravitz, J. (2002, June). *Demystifying data about technology impacts in schools.* Paper presented at the National Educational Computing Conference (NECC). San Antonio, TX. [WWW Document]. Available: http://www.bie.org/Ravitz/Ravitz_NECC_2002.pdf

Ravitz, J. (1999). *After the wires: Implementation conditions related to increased teacher use of the Internet in schools with high Internet connectivity*. Paper presented at Annual Meetings of the American Educational Research Association. Montreal, Canada. [WWW Document]. Available: http://www.bie.org/Ravitz/Ravitz_AERA_1999.pdf

Ravitz, J. & Mergendoller, J. (2003, forthcoming). Triangulating surveys and observation methods to assess professional development impacts: Focusing on the interaction of beliefs about good teaching and the development of technology skills. In M. Russell (Chair), *Technology research symposium: A focus on four under-used methodologies*. Paper to be presented at Annual Meetings of the American Educational Research Association. Chicago, IL.

Ravitz, J., Becker, H. & Wong, Y. (2000*). Constructivist compatible beliefs and practices among U.S. teachers. Report #4. Teaching Learning and Computing: 1998*. Irvine, CA: University of California at Irvine, Center for Research in Information Technology in Organizations. [WWW Document]. Available: http://www.crito.uci.edu/tlc/findings/report4

Ravitz, J., Mergendoller, J. & Rush, W. (2002). *What's school got to do with it? Cautionary tales about correlations between student computer use and academic achievement.* Paper presented at Annual Meetings of the American Educational Research Association. New Orleans, LA. [WWW Document]. Available: http://www.bie.org/Ravitz/Ravitz_AERA_2002.pdf

Rockwell, S. & Kohn, H. (1989). Post-Then-Pre Evaluation. *Journal of Extension, 27*(2). [WWW Document]. Available: http://www.joe.org/joe/1989summer/a5.html

Rogers, E. (1983). *Diffusion of Innovations. (3rd ed.)*. New York: Free Press.

Thompson, B. (1999). Five methodology errors in educational research: A pantheon of statistical significance and other faux pas. In Bruce Thompson (Ed.). *Advances in social science methodology: Volume 5*. Stamford, CT: JAI Press, pp. 23-86.