Educational Research Journal《教育研究學報》*, Vol. 25, No. 1, Summer 2010*

What Are the Chances of Success for My Project? And, What If It Was Already Done? Using Meta-Analyzed Effect Sizes to Inform Project Decision-Making

Kay-cheng Soh

Singapore

Action research projects have been used as a means of innovations in curriculum and in-service professional up-grading of teachers. Many instructional innovations of the Western world are being replicated in Asian's education systems often without due consideration of success rates. As such endeavour consumes much time and other resources, it would be prudent that account be taken of success rates based on the Western experiences. This paper suggests using meta-analyzed effect sizes hitherto available for (1) making pre-project decisions as to whether a project is worth trying in view of the relevant past experiences, and (2) evaluating a project effect, post hoc, vis-a-vis the average effect size available for similar projects for a contextualized interpretation. Examples and information for doing these are given. Relevant conceptual issues are discussed.

Key words: meta-analysis, action research, curriculum innovation

---

Correspondence concerning this article should be addressed to Kay-cheng Soh, 50 Lorong 40 Geylang #07-29, The Sunny spring, Singapore 398074. E-mail: sohkc@singnet.comn.sg

Those who cannot remember the past are condemned to repeat it.

Santayana (1905)

Asian nations such as Hong Kong, Korea, Singapore, Thailand, and Vietnam are going through a phase of education reform in the recent past. This entails re-designing the curriculum and introducing instructional innovations. In this effort, many supposedly innovative ideas which have been tried out in the Western nations, especially the United States, with varying degree of success, are replicated with the belief that they help. However, collated evidence of the efficacies of these attempts is yet to be procured.

Innovative ideas tried out include problem-based learning, inter-disciplinary teaching, multiple intelligences approach, understanding by design, individualized instruction, differentiated teaching, mind mapping, self-directed learning, Da Vincian Principles, *La main a la pate* (The hands with the paste), Socratic Questioning, journaling, peer tutoring, etc., etc. The list continues. And, of course, we do not forget the ubiquitous ICT which includes a long list of electronic devices and approaches.

Action research of one kind or another has been used, first, as a way of finding out the efficacy of a wide range of so-called innovative measures to improve teaching effectiveness. Action research is also used as a way of teachers' in-service professional up-grading. For instance, a very large number of teachers have been trained though school-based or zonal workshops to conduct action research projects and by a programme at the system level for training "research activists" (Soh, 2006). The dual objective of the effort is to encourage and equip teachers for conducting school-based curriculum innovation projects and to improve the quality of teachers in their professional skills.

As education reform especially innovations in instruction is resource-intensive and time-consuming and, often, controversial, it pays to ask a fundamental question before any effort to change is embarked on. Very often, with the enthusiasm, such projects are undertaken without sufficient consideration for what has been found in the pertinent literature. These give rise to the question: *What are the chances of success?* After all, if you don't try, you won't be able to tell. It is such faith and not fact that drives the innovation efforts. As aptly put by Hattie (1999),

> Hence, we have a school community peopled with teachers with self-fulfilling prophecies, all believing they are doing a good job, and with models of learning rarely based on any other evidence than "it works for me."

Since action research projects have been undertaken without much consideration for their chances of success, the next question naturally arises: *What if it's already done?*

Conventionally, project efficacy has been evaluated by the ubiquitous independent *t* test. However, the relevance of this and similar statistical significance tests have been called to question, especially in the context of school-based quantitative action research. Two separate but related issues are involved here. First, the *t* test result (and its corresponding *p* value) answers the question "*What is the chance of getting a group difference of the magnitude obtained or greater, if the null hypothesis is true in the population*?" In this case, a project may produce a statistically significant difference which may be too small to be of practical or educational significance or importance (Kirk, 1996). And, as is well-known, the statistical significance of a *t* value is dependent on the sample size such that a small *t* value can be statistically significant when the sample size is large (Kirk, 2001). In contrast, action researchers are concerned with whether the intervention (action) has produced the expected effect and if so how large it is. This is a question that can only be answered by the effect size (Soh, 2008, 2009b) but not the probability of chance occurrence which the *p*-value answers (Glass, McGaw, & Smith, 1981; Thomson, 1998).

Secondly, action researchers' primary concern lies with finding solutions to practical problems in the school or classroom context and not with generalizing the results to assumed populations. In fact, for action research, it is doubtful if there is a population in the strict sense of the terms at all (Soh, 2009a). It is with such realization that effect size indicators have become important to quantitative research such that many journals now require reporting effect sizes in addition to significance test results (Coe, 2002).

This paper is an effort to provide school administrators and teacher-researchers with a conceptual framework and relevant information available so far to help them make decision before and after the conduct of their action research projects.

What Are the Chances of Success for My Project?

Effect size is typically used at the conclusion of a research project to ascertain its success or the lack of it. However, with a very large number of meta-analyses (more than 800 covering 165,258 studies) available where more than 130 different "influences" covering such domains as student, teacher, teaching, curricula, home, school (Hattie, 2009), it would be wise to consult them before making a decision whether to proceed with an intended project. Doing this will enhance the understanding of what may succeed and what may not even before research effort and resources are deployed. Although there may be a risk of foreclosing a project because previous similar projects have shown small or even negative effect sizes.

Let's take a few examples to show how such meta-analyzed project effect sizes can be used to guide decision-making. In Table 1, the first column shows the kind of innovations tried to enhance student achievement, the numbers of relevant studies are shown in the next column. The third column shows the average effect sizes for studies of a particular innovation effort, followed by the standard deviation (*SD*) for those studies. The last two columns are the corresponding estimated lower and upper limits for 95% confidence.

The limits are calculated by subtracting from the average effect size 1.96 times of the *SD* for the lower limit and the average effect size plus 1.96 times of the *SD* for the upper limit. Assuming normal distribution of the effect sizes of the relevant studies, the effect size of a new similar project is likely to be found within the limits, with 95% certainty.

Take for example the case of reinforcement for which students are given positive feedback on their learning performance — encouraging students for doing the right things. The average effect size of the 139 studies using reinforcement is 1.13. It has a *SD* of 0.12 and hence a lower limit of 0.89 and a upper limit of 1.37. This suggests that if a project is planned to use reinforcement to enhance student achievement, it will produce an effect of the size at least 0.89 which is, again, a large one by Cohen's criterion. Even more optimistically, the effect size may be as high as 1.37. In short, such a project is destined to be highly effective.

Table 1: Selected Effect Sizes of Various Innovation Efforts

Innovation	No. of studies	Effect size	*SD*	Lower limit	Upper limit
Reinforcement	139	1.13	0.12	0.89	1.37
Class environment	921	0.56	0.36	–0.15	1.27
Simulation and games	111	0.34	0.01	0.32	0.36
Computer-assisted instruction	566	0.31	0.14	0.04	0.58
Individualization	630	0.14	0.14	–0.13	0.41
Physical attributes of school	1850	–0.05	0.17	–0.38	0.28

Take the second example of classroom environment. This may include making the classroom physically attractive or ensuring psychologically conducive for learning. As shown in Table 1, the 921 studies in this innovation effort have an average effect size of 0.56 with a *SD* of 0.36. Compared with the reinforcement case, this effect size is much smaller, in fact, in the range of medium effect according to Cohen's criterion. The larger *SD* suggests that the effect is comparatively less certain. The lower limit of –0.15 suggests that a new similar project may have a negative effect on student achievement (!) or a neutral effect as the limits include zero. But, the large positive upper limit suggests that it may also produce a large desired effect. Since the limits cover more positive values, say, from 0.20 (for small effect) up to 1.27 (for large effect), the project has more chances for a success. It is therefore worthy of investing effort and resources in it.

Simulation and games as a method to enhance student achievement has been popular with teachers for some time and supposedly continues to be popular. As the third example, it has a small averaged effect size of 0.34 and a rather small *SD* of 0.01. This suggests that it is rather certain that such a project will produce a small effect size. Thus, a new project may have an effect size as low as 0.32 or as high as 0.36, both falling within Cohen's small effect size range.

Computer-assisted instruction may take various formats and the most common one is the drills-and-practice lesson (which is a modern version of programmed instruction). This is not a time to trace its popularity but obviously a large number of studies in this regard have been conducted, 566 to be specific. The average effect size for these is 0.31 (*SD* = 0.14). This being the case, the lower limit is 0.04 and upper limit 0.58, suggesting that a new project in this line may produce negligible or low effect (0.20 or below) but more likely a small effect (effect size up to 0.58). This may sound somewhat disappointing in view of the large amount of financial resources invested in equipping the classroom and the faith (somewhat blindly?) in anything computer. However, the meta-analyzed result is a fact to be accepted or at least tolerated.

The fifth example is individualization. Again, this innovation effort may take many forms but the basic idea is to attend to the learning needs and styles of individual students. The assumption is that if teachers spend more time with the individuals (not as a class), their learning will be more effective. The fact is, based on the large number of studies (630), the effect does not seem to support it with an average effect size of 0.14 (trivial, according to Cohen). With a *SD* of 0.14, there is a possibility of low negative or null effect, although there is a possibility of a low positive effect. In short, a new project following this line of thinking may not be as encouraging as it seems reasonable.

The sixth and last example comes from a very large number of studies (1850) of the effect on student achievement of physical attributes of school. Physical attributes of schools is the very first contact point between the community and its education system, for the simple reason that the school campus is what everyone can see while what goes in the name of education is more elusive to laymen. It is therefore natural for education authorities the world over to spend a very amount of their financial resources to ensure school's physical attributes. The sad fact from the meta-analysis is that the average effect size is –0.05 (0.17). Thus, it is almost certain that spending money on school's physical attributes is not going to pay off, what a lower limit of –0.38 and a positive upper limit of 0.28. If this experience of a very large number of studies is something to go by, the money may produce better return of investment (ROI) if used for other more effective innovation efforts (those higher in the table appended).

The above examples illustrate how the meta-analyzed effect sizes can inform decision-making for school-based and classroom-based action research projects. Basically, we look first at the average effect size for a particular kind of projects of interest to see what is the likely magnitude of project effect, and then at the lower and upper limits (which are based on the *SD*) to assess the chance of success; those covering 0.5 (medium) and above are more likely to be successful where student achievement is concerned.

Armed with the concept and technique exampled above, therefore, it is a pertinent question to assess the likelihood of success when planning a curriculum innovation action research project. As such projects use up much time, effort and resources, it is natural that responsible school administrators and teacher-researchers ought to be concerned with the question.

However, as an answer to the question is not always available, many curriculum innovation projects are decided upon based on personal preferences, faith, or beliefs rather than on objective evaluation based on past experience. While preferences, faith, and beliefs may be necessary to motivate the projects, they should not be the only reasons that a project is put in place.

At the beginning of a project, school administrators and teacher-researchers concerned will be well-advised to secure relevant information to help in evaluating the likelihood of project success so as to avoid trying something they consider innovative but past research has shown little promise. Of course, for action research projects, teacher-researchers usually review about a score of relevant papers, and even doing this *post hoc*. This tends to be selective rather than comprehensive and thus becomes self-serving.

There is in fact ample relevant information to guide the decision. The table in the Appendix re-produces meta-analyses of 165,258 studies of innovation efforts conducted in the Western world, mainly America, prior to 1999 (Hattie, 1992, 1999, 2009). The average effect size is 0.40

which is small but near medium in magnitude according to the criterion recommended by Cohen (1988). Hattie (2009) recommends that this effect size of 0.4 be used as a benchmark, in his “barometer” for innovations.

Needless to say, some of the effect sizes are far above this average and other far below. It, therefore, would be wise to replicate studies which have been found to be successful in such a large number of previous studies and to avoid those otherwise. This pragmatic approach is vital to decision-oriented action research which aims at solving practical curricular and instructional problems, as contrasted with academic research which aims at enhancing understanding and hence conclusion-oriented.

By way of summary, the meta-analyzed average effect sizes help school administrators and teacher-researchers when planning projects to see the chances of success of their curriculum innovation efforts, and the standard deviations (*SD*s) suggest the certainty of that likelihood. In other words, the information can be used for preliminary evaluation of whether a project being considered is likely to be successful and hence worthy of efforts and resources.

What If It Was Already Done?

If an action research project has been conducted without due consideration for success, this is a question a school administrator and teacher-researchers will naturally ask at the completion of a curriculum innovation action research project.

Theoretically speaking, there are three possible outcomes of any such project, a positive one, a neutral one, and a negative one. The negative one is disappointing, the neutral one puzzling, and the positive one encouraging. Even if the outcome is positive, it may not be large enough to satisfy the school administrator and teacher-researchers concerned. This could be due to the time, effort, and resources incurred and, perhaps, the desire to impress.

In sum, the obtained effect size of an action research project in curricular and instructional innovation has to be evaluated with reference to what has been reported in the pertinent literature and not only by the absolute criterion (such as that recommended by Cohen) or subjective expectation. This objective approach to evaluation of project outcome helps to minimize the ill-effect of wishful thinking.

Efforts to bring about enhanced student achievement through curricular and instructional innovations have been there for the past few decades. It will continue into the future as nations the world over strive for excellence. Hattie's (1992, 1999, 2009) meta-analysis covers a very large number of studies of varied approaches from reinforcement, through homework and instructional media, to behavioral objectives. It thus gives a trustworthy and comprehensive view of what works and what does not in curricular and instructional innovations.

For these meta-analyses, the average effect is 0.40 (0.19). This indicates a lower limit of 0.03 (trivial effect) and upper limit of 0.77 (medium but nearly large effect). Lest this becomes discouraging to school administrators and teacher-researchers, a few reservations are necessary to put the situation in its right perspective.

Firstly, as alluded to above, student achievement is influenced by a host of factors some of which are beyond research intervention (e.g., student ability, home support). Thus, any action research project can only manipulate one or two such factors as its focus, leaving all the rest as confounding or uncontrolled factors. This is the main reason why project effects have not been more "impressive" than being medium on average.

Secondly, many of the methodological details and differences of the studies are not visible in Hattie's papers due to the meta-analysis process. This is rightly so as the aim of a meta-analysis is to summarize the findings of a large number of studies without burdening the readers with a lot of details peculiar to specific studies.

Thirdly, it may be argued that the studies were conducted in the Western context and the findings may not translate in an Eastern culture. While there is some degree of validity in this, empirical evidence is being accumulated by cross-cultural researchers and it is premature to conclude one way or the other (e.g., Li, 2003; Watkins, 2000).

Fourthly, and most critical, Hattie's meta-analysis was completed in and around the nineties. After its publication, more than a decade has passed since and there must be new studies which confirm or contradict.

Until a revised meta-analysis is done, for the time being, Hattie's summary seems to be the only information available to guide decision-making. As those summarized studies span over a long period of time with a large number in many locations, things might not have changed much.

Caveats

Admittedly, decisions on whether to proceed with school- or classroom-based action researcher projects involve more than just the probability of their success. Socially and culturally relevant considerations have their role to play in the decision-making process. Moreover, an action research project may have more than one main concern in terms of a single effect size. Therefore, averaged effect sizes found in the literature for various kinds of projects should not dictate but inform project decision-making. Nevertheless, when producing the desired outcomes in terms of student achievements is the major if not the only project goal, then there is nothing to lose but everything to gain by consulting the relevant meta-analysis prior to deciding to proceed as planned.

A few big-bang projects involving a large sample may be more impressive and sensational, but the principle of converging evidence is more relevant, since a case does not make a rule, especially in education which is a highly complex endeavour on changing human (student) thinking, feeling, and behaving. In this regard, school administrators and teacher-researchers will be well-advised to take note of what Stanovich (2001) has to say,

> The reason for stressing the importance of convergence is that conclusions in psychology (read *education*) are often based on the principle of converging evidence. There is certainly nothing unique about this fact. Conclusions in many other sciences rest not on single, definitive experimental proofs, but on the confluences of dozens of fuzzy experiments…. Experiments…are usually of fairly low dignosticity. That is, the data that support a given theory usually rule out only a small set of alternative explanations, leaving many additional theories as viable candidates. As a result, strong conclusions are usually possible only after data from a very large number of studies have been collected and compared. (p. 130)

References

Coe, R. (2002, September). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of education*, *36*(1), 5–13.

Hattie, J. (1999). *Influence on student learning*. Professor of Education Inauguarl Lecture, University of Auckland, New Zealand.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London, England: Routledge.

Kirk, E. R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746–758.

Kirk, E. R. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, *61*, 213–218.

Li, J. (2003). U.S. and Chinese cultural beliefs about learning. *Journal of Educational Psychology*, *95*, 258–267.

Soh, K. C. (2006). Promoting action research in Singapore schools. *New Horizons in Education*, *53*, 8–21.

Soh, K. C. (2008). Effect size: What does it do for educational action researchers? *North Star: A Publication for Educational Practitioners*, *1*(1), 63–70.

Soh, K. C. (2009a). *Analyzing data and interpreting outcomes: Statistical toolbox for teacher-researchers*. Singapore: Cobee Publishing House.

Soh, K. C. (2009b). Effect size (again): The point-biserial correlation. *North Star: A Publication for Educational Practitioners, 1*(2), 58–64.

Stanovich, K. (2001). *How to think straight about psychology* (6th ed.). Boston, MA: Allyn and Bacon.

Thomson, B. (1998). Statistical significance and effect size reporting: Portraits of a possible future. *Research in Schools*, *5*, 33–38.

Watkins, D. (2000). Learning and teaching: A cross-cultural perspective. *School Leadership and Management*, *20*(2), 161–173.

Appendix: Effect Sizes of Various Innovation Efforts

Innovation	No. of studies	Effect size	*SD*	Lower limit	Upper limit
Above-average effect size (0.40) for 165,258 studies					
Reinforcement	139	1.13	0.12	0.89	1.37
Cognitive ability of pupils	896	1.04	0.36	0.33	1.75
Quality of instruction	22	1.00	0.00	1.00	1.00
Quantity of instruction	80	0.84	0.03	0.78	0.90
Direct instruction	253	0.82	-	-	-
Acceleration	162	0.72	-	-	-
Home support	728	0.67	0.17	0.34	1.00
Disposition to learn of pupils	35	0.61	0.08	0.45	0.77
Remediation/feedback	146	0.65	-	-	-
Background of instructor	93	0.60	0.00	0.60	0.60
Class environment	921	0.56	0.36	–0.15	1.27
Challenge of goals	2703	0.52	-	-	-
Bilingual programmes	285	0.51	-	-	-
Methods (Reading)	557	0.50	0.29	–0.07	1.07
Tutoring	125	0.50	0.12	0.26	0.74
Mastery learning	104	0.50	0.08	0.34	0.66
Teacher in-service education	3912	0.49	-	-	-
Parental involvement	339	0.46	-	-	-
Homework	110	0.43	-	-	-
Style of instructor	264	0.42	0.28	–0.13	0.97
Questioning	134	0.41	-	-	-
Below-average effect size (0.40) for 165,258 studies					
Peer influence	122	0.38	0.00	0.38	0.38
Advance organizers	387	0.37	0.25	–0.12	0.86
Methods (Science)	730	0.34	0.12	0.10	0.58
Simulation and games	111	0.34	0.01	0.32	0.36
Computer-assisted instruction	566	0.31	0.14	0.04	0.58
Methods (Mathematics)	416	0.30	0.36	–0.41	1.01

Instructional media	4421	0.30	0.25	–0.19	0.79
Methods (Others)	60	0.28	0.00	0.28	0.28
Aims and policy of school	542	0.24	0.31	–0.37	0.85
Affective attributes of pupils	355	0.24	0.12	0.00	0.48
Calculators	231	0.24	-	-	-
Physical condition of pupils	905	0.21	0.14	–0.06	0.48
Learning hierarchies	24	0.19	0.00	0.19	0.19
Ability grouping	3385	0.18	-	-	-
Programmed instruction	220	0.18	0.10	–0.02	0.38
Audio-visual aids	6060	0.16	-	-	-
Individualization	630	0.14	0.14	–0.13	0.41
Behavioral objectives	111	0.12	0.00	0.12	0.12
Finances/money	658	0.12	-	-	-
Team teaching	41	0.06	0.00	0.06	0.06
Physical attributes of school	1850	–0.05	0.17	–0.38	0.28
Mass media influence	274	–0.12	0.00	–0.12	–0.12
Retention	861	–0.15	-	-	-

Note: Adapted from Hattie (1992, 1999). For more information, see Hattie (2009, Appendix B). Where *SD* is not available, the lower and upper limits are not estimated.