What are the chances of success for my project?
And, What If It Was Already Done?
Using Meta-analyzed Effect Sizes for Project Decision-making

Soh, Kay Cheng

Abstract

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 experience and (2) evaluating a project effect, *post hoc*, vis-a-vis the average effect size available for similar projects for a contextualized interpretation. Examples for doing these are given. Relevant conceptual issues are discussed.

Those who cannot remember the past are condemned to repeat it.
Santayana (1905)

Asian nation 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 USA, 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 teacher's 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 do a good job, and with models of learning rarely based on any other evidence than "it works for me".

What are the chances of success for my project?

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?*

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?

This is a pertinent question to ask 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 he 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.

What are the chances of success for my project?

There is in fact ample relevant information to guide the decision. Table 1 integrates 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.

Table 1 about here

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.

Let's take one or two examples. For instance, team teaching was in vogue some decades ago on the ground that it made good use of the team members' knowledge, saved their preparation time, and maximized their respective strengths in instruction. This sounds reasonable. The disappointing fact is meta-analysis shows the effect size for this innovation to be near zero (0.06) and hence trivial, by Jacob Cohen's criterion (Cohen, 1997; Soh, 2008, 2009). With a standard deviation (SD) of zero, these projects invariably were unable to produce the expected benefits! This being the case, it needs a lot of courage and faith to undertake a project in the face of such negative evidence.

On the other hand, meta-analysis of studies of quality of instruction yields an effect size of 1.00 which is large by Cohen's criterion. This suggests that to enhance student achievement, ensuring quality instruction is a sure bet. Moreover, the zero SD indicates that all the project meta-analyzed have produced the expected effect of enhanced achievement. This underlines the importance of good teachers and the need to ensure teachers are good (however this is defined). As a matter of fact, Hattie's (20XX) meta-analysis shows that teacher accounts for around 30% of achievement differences among students, second only to student which accounts for 50% of the differences.

Things may not always be as clear as the two cases. For example, class environment has an average effect size of 0.56 (0.36). This effect size is medium for the 102 studies meta-analyzed. This suggests that efforts to make class environment physically attractive and psychologically conducive may bring about a benefit in student achievement neither large nor small but definitive. However, the SD of 0.36 suggests that about two-thirds of the studies had effect sizes varying from a low 0.20 (0.56 – 0.36; small effect) to 0.92 (0.56 + 0.36; large effect). Thus, enhancing class environment may produce from small to large effect with less certainty; other factors come into play, too.

Take yet another example, that of simulation and games. The effect size of 0.34 (0.01) indicates that these learning activities had a small effect on student achievement. Its certainty is suggested by the almost near zero SD of 0.01. In other words, the one-SD bounds are 0.33 to 0.35 and both are small by Cohen's criterion. This suggests that a project of simulation and games has a two-in-three chance of improving student achievement by about one-third of a SD, small by Cohen's criterion. With reference to the normal curve, if a group of students not using stimulation and games are considered as standing at the 50th percentile (average), the a group of students using stimulation and games will stand at the 63rd percentile, that is, 13 positions ahead.

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 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.

For instance, a school has spent a large amount of fund for equipment and training to mount an ICT-based curriculum. It is natural that a large expenditure to equip the classroom and the teachers lead to a high expectation. If the effect size turns out to be

0.38, this small though positive outcome may be disappointing, in terms of ROI (return of investment). But, it need not be, because an effect size of this magnitude is close to but somewhat greater than the meta-analyzed value of 0.31 (0.14) as shown in Table 1 for computer-assisted instruction. In other words, the outcome is well within the range of 0.17 (0.31 – 0.14; small effect) to 0.45 (0.31 + 0.14; medium effect) for two-thirds of the 557 studies meta-analyzed. The teacher-researchers can legitimately conclude that the project has produced an effect comparable to that found in the international scene and an expectation for larger effects may be unrealistic. This also serves to remind that student achievement is multiply determined by a host of factors of which ICT-based instruction is but one.

Take another example, that of individualization. If a school, for some reasons, decides to conduct an action research project of individualized instruction in spite of the meta-analyzed effect size of 0.14 (0.14) as shown in Table 1, then, it ought to be prepared for an outcome between 0.00 (no effect) and 0.28 (small effect), since two-thirds of the studies fall within this range of effect size. Had the obtained effect size be 0.42, the project is placed at the top of the 98% (0.14 + 2*0.14; large effect) of the very large number of studies. Although 0.42 is still an effect which is small though nearing medium (0.50), something unusual must have happened. This being the case, the teacher-researchers should try to find out what these could be, besides joining in celebrating the unusual achievement.

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.

Caveats

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, 20XX) 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.

The average effect is 0.40 (0.19). This indicates that two-thirds of the studies have effect sizes varying from small (0.21) to medium (0.59). Lest this becomes discouraging to school administrators and teacher-researchers, a few caveats 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.

Conclusion

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 -- Soh) 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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd edition*. Hillsdale, NJ: Erlbaum.

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 Inaugular Lecture, University of Auckland, New Zealand.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-analysis Relating to Achievement.* London: Routledge.

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. (2009). 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 edition*. Boston: Allyn and Bacon.

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

Table 1. Effect sizes of various innovation efforts

Innovation	No. of studies	Effect size	SD
Above-average effect (0.40 for 165,258 effect sizes)			
Reinforcement	139	1.13	0.12
Cognitive ability of pupils	896	1.04	0.36
Quality of instruction	22	1.00	0.00
Quantity of instruction	80	0.84	0.03
Direct instruction	253	0.82	-
Acceleration	162	0.72	-
Medium effect			
Home support	728	0.67	0.17
Disposition to learn of pupils	35	0.61	0.08
Remediation/feedback	146	0.65	-
Background of instructor	93	0.60	0.00
Class environment	921	0.56	0.36
Challenge of goals	2703	0.52	-
Bilingual programmes	285	0.51	-
Methods (Reading)	557	0.50	0.29
Tutoring	125	0.50	0.12
Mastery learning	104	0.50	0.08
Teacher in-service education	3912	0.49	-
Parental involvement	339	0.46	-
Homework	110	0.43	-
Style of instructor	264	0.42	0.28
Questioning	134	0.41	-
Below-average effect			
Peer influence	122	0.38	0.00
Advance organizers	387	0.37	0.25
Methods (Science)	730	0.34	0.12
Simulation and games	111	0.34	0.01
Computer-assisted instruction	566	0.31	0.14
Methods (Mathematics)	416	0.30	0.36
Instructional media	4421	0.30	0.25
Methods (Others)	60	0.28	0.00
Aims and policy of school	542	0.24	0.31
Affective attributes of pupils	355	0.24	0.12
Calculators	231	0.24	-
Physical condition of pupils	905	0.21	0.14
Trivial effect			
Learning hierarchies	24	0.19	0.00
Ability grouping	3385	0.18	-
Programmed instruction	220	0.18	0.10
Audio-visual aids	6060	0.16	-

Individualization	630	0.14	0.14
Behavioral objectives	111	0.12	0.00
Finances/money	658	0.12	-
Team teaching	41	0.06	0.00
Physical attributes of school	1850	-0.05	0.17
Mass media influence	274	-0.12	0.00
Retention	861	-0.15	-

Note: Adapted from Hattie (1992, 1999). For more information, see Hattie (2009, Appendix B).