Types of Errors in Synthesizing Research in Education

Michael J. Dunkin
The University of New South Wales

Nine types of errors occurring in three stages of the process of synthesizing research are described and illustrated with examples from a recent synthesis (Kagan, 1992) of research on teacher professional growth. Errors can occur in the initial identification and collection of reports of research, in the analysis of documents, and in the final stage of reaching generalizations about the whole body of research. Types of errors include the exclusion of relevant literature, wrongly reporting details such as sample size, erroneously attributing findings to studies, and stating unwarranted conclusions about the research reviewed. Implications for reviewers and users of reviews are considered.

In the last 20 years a large body of literature on ways of synthesizing research in education has developed (Dunkin, 1994; Walberg, 1986). Approaches to gleaning the accumulated findings of that research have varied from the narrative through vote counting or box scores to meta-analysis. Some of these approaches make more demands on the conceptual and interpretative skills of the synthesizer than others and, therefore, contain more scope for error and bias than others, although all approaches are subject to the fallibility of the synthesizers and those upon whom they necessarily rely. It is important that the validity of all syntheses be tested, for they are the main ways in which assessments can be made about the accumulation and development of research-based knowledge.

Syntheses of research are influential in regard to subsequent research, policy, and practice. They provide the empirical bases for applications for research grants, for higher-degree dissertations and theses, and for individual and institutional research. They are used by policymakers in designing strategies for development, and they are used to guide practitioners in the enhancement of professional activity. They provide the contents of highly regarded publications in handbooks, encyclopedias, and textbooks and become the best known statements of the state of knowledge on the topics to which they are addressed.

The processes by which syntheses of research are conducted and disseminated are, therefore, crucially important, because they determine the quality of the syntheses and which syntheses are available publicly for the above purposes. If the synthesizers allow systematic biases to affect their selection of studies to review,

The author would like to thank the anonymous referees and the editors for their helpful advice in bringing this paper to the state of publishability. They are not in any way responsible for any errors that might be found in it.
Dunkin

if they fail to recognize that some authors of the original sources were wrong in announcing their findings, if they allow their own priorities to affect the findings they report to the exclusion of contrary findings, then there is the potential for a synthesis that is seriously flawed. The likelihood that a poor synthesis would survive the rigorous refereeing process employed by prestigious scholarly journals is undoubtedly very small. Nevertheless, the consequences of such a mistake warrant contemplation, for they could be whole programs of misguided research, policy, and practice.

The purposes of this article are to suggest the stages at which a synthesizer is at risk of making mistakes, to present a typology of errors that can be made, and to illustrate the typology with errors made in a recent synthesis—namely, Kagan’s (1992) review of studies on professional development among preservice and beginning teachers.

Stages in the Occurrence of Error

There are three stages at which synthesizers might make errors. The first, the primary stage, is when the synthesizer searches the literature and selects from it the items judged relevant to the topic of the review. Errors made at this stage result in bias that might lead to conclusions that represent the findings of only part of the research and omit the findings of the rest, or that give equal status to the findings of good and poor research.

At the secondary stage, the reviewer analyzes the literature selected in order to identify context, methods, and the findings of each study included. This is the stage at which the variety of errors made is greatest. As detailed below, error in identifying facts about contexts and methods leads to the misclassification of studies, and errors in identifying and reporting findings introduce error into the next stage of the synthesis.

This next stage, the tertiary stage, occurs when the synthesizer accumulates the findings identified in the previous stage in order to reach generalizations about the topic under investigation. Errors brought forward from the primary and secondary stages have their fullest impact at this stage, for they can lead to invalid generalizations.

Types of Reviewer Error

Primary Stage Errors

Type 1: Unexplained selectivity. These are errors in which the reviewer excludes research which comes within the declared scope of the review without explaining or justifying the exclusion. The outcome of this error is that the conclusions of the reviewer cannot be held to apply to the whole defined field of concern. It is possible, for example, that the findings of the body of the excluded research contradict those of the included research, and that reliance on the latter alone produces a biased picture of the state of knowledge in the field. Contemporary facilities for conducting literature searches and developing bibliographies are so advanced that these errors can seldom be explained in terms of understandable reviewer ignorance of the existence of the excluded work.

In her critique of Kagan’s (1992) review, Grossman (1992) specified 16 studies that came within the declared boundaries of the review but had been ignored. She
Errors in Synthesizing Research

also drew attention to the fact that a surprisingly high proportion of the papers analyzed had been written by just two authors, which suggests that relevant papers by other authors had been ignored. If Grossman was right, then the review cannot be regarded as adequately representing the state of research knowledge on teacher professional growth, because of the presence of Type 1 errors.

Type 2: Lack of discrimination. Not all research on the same topic is of equal quality. Much of the literature available to a reviewer consists of tentative reports in the form of conference papers presented by authors seeking feedback prior to preparation of final drafts, which may or may not be submitted for publication in refereed scholarly journals. One test, therefore, of the quality of a conference paper is whether or not it subsequently appears in one version or another, perhaps under a different title, in a refereed journal. In their efforts to be up-to-date, reviewers often do not wait that long, and rely on conference papers themselves. Clearly, this is a risky practice which can result in preliminary findings being accorded the same status as findings contained in journal articles that have survived rigorous refereeing and editing processes. This is not to say that all conference papers are defective or that all journal articles are free of weaknesses. In either case, reviewers should be vigilant that they distinguish between good and poor research, lest they give equal status to both.

Grossman (1992) accused Kagan (1992) of having committed this type of error but did not specify examples of poor research reviewed. This is possibly the most difficult type of error to identify, because criteria for evaluating research are usually controversial, and judges sometimes disagree on the difference between good and poor research. Most of the studies included in Kagan’s review had been published in refereed journals, but 11 were cited as conference papers presented at Annual Meetings of the American Educational Research Association. Three of those papers could not be obtained for present purposes. Hollingsworth (personal communication, September 29, 1994) reported that she did not coauthor the paper cited as “Lidstone and Hollingsworth (1990),” and attempts to contact the other cited author were unsuccessful. The paper was not listed in the main, printed program of the meeting at which it was said to be presented, and it seems not to have been published elsewhere. The paper cited as “Wendel (1989)” was also unavailable for this analysis. It did not appear in the main, printed program of the meeting concerned, and repeated attempts to locate the paper were unsuccessful. Finally, the paper by Cochran-Smith (1989) was revised and subsequently published (Cochran-Smith, 1991) under a different title. During the years covered by Kagan’s review (1987 through 1991), the present author was the editor of Teaching and Teacher Education: An International Journal of Research and Studies. Three of the papers reviewed by Kagan were submitted for publication in that journal and were subsequently rejected.

The unavailability and/or nonpublication of some papers in their original form, the form which was used in Kagan (1992), may well indicate something about their quality and give justification to Grossman’s (1992) criticism.

Secondary Stage Errors

As mentioned above, Grossman (1992) criticized Kagan (1992) mainly for making two types of errors, which have been classified as primary stage errors here. She said little, however, about the types of errors described below that also
were found in Kagan’s synthesis.

**Type 3: Erroneous detailing.** These errors consist of incorrect statements of the sampling, methods, designs, procedures, and contexts of the studies reviewed. In some cases, the reports reviewed are about parts of larger projects and contain descriptions of the larger projects as well as the relevant parts. Details of the whole are sometimes thought to apply to the part when, in fact, they do not. In other cases, the reviewer accepts an opening statement of sample size without recognizing that attrition occurred and that the data actually came from a smaller sample. Errors in the reporting of methods, procedures, designs, and contexts can lead to the misclassification of studies so that they do not share the essential characteristics that are supposed to make them comparable with other studies.

In Kagan (1992) the sample sizes of four studies were reported inaccurately. In one case (Calderhead & Robson, 1991) the sample size was said to be 12, when it was 7; in another (McDaniel, 1991) it was said to be 22, when it was 3; in a third study (Borko et al., 1991), it was said to be 38, when it was just 1; and in the fourth (Wildman, Niles, Magliaro, & McLaughlin, 1989) it was said to be 15, with in-depth profiles of 4, when in fact the 15 were the subject of an earlier report (Wildman, Magliaro, Niles, & McLaughlin, 1988), and only the 4 case studies were reported in the publication reviewed. The main problem in these cases was that the papers reviewed were sometimes reports of parts of larger studies. The larger sample sizes were true of the full-scale studies, but the parts of these studies being dealt with in the review had smaller sample sizes. Thus, all four errors were in the direction of larger sample size, so that the reader was led to believe that the accumulated findings of those four studies came from a total of 87 novice teachers, when in fact they came from a total of just 15!

Other Type 3 errors were made in reporting data gathering procedures. For example, concerning the study by Aitken & Mildon (1991), it was claimed in Kagan (1992) that only interviews were used, when in fact there were workshop presentations, written preparations for these workshop presentations, self-evaluations, peer evaluations, written autobiographies, written responses to workshops, and researchers’ field notes.

A more serious error concerning these types of details led to the complete misclassification of three studies. The Laboskey (1991) study was classified in Kagan (1992) as being concerned with the “image of self as teacher “ (p. 146). Laboskey made no reference to self-image, and nowhere in the paper was anything said about the “central role” that self-image was said by the reviewer to play. In the section headed “Requisites for Growth During Practica and Student Teaching,” the review summarized six studies which, it was claimed, “examined how candidates’ knowledge of teaching changed during a practicum, student teaching, or the course of an entire preservice program” (p. 140). In fact, one of the studies (Gore & Zeichner, 1991) did not investigate change at all. Then, the study by Chamberlin & Vallance (1991), included in the group supposed to be “comprehensive evaluations of practica or student teaching experiences,” was said to involve student teachers in spending half of each day in a 9-hour “block course” in classrooms. In fact, the student teachers in this study spent only half a day per week for 9 weeks in classrooms. The other studies reviewed in that section involved either 1 or 2 semesters or 1 year of full-time student teaching and were, therefore, clearly not comparable to the misclassified study. Needless to say, the
Errors in Synthesizing Research

findings of that study could not legitimately be claimed to contribute to any accumulation of evidence that might be perceived in the rest of the group.

Type 4: Double counting. This error consists of listing different reports from the same project as providing additional confirmation of the same finding. The risk of this error occurring seems to be present particularly when there is multiple reporting of results of the one project. This can be difficult to detect, especially given that titles and lists of authors are sometimes changed. While it might be expected that these multiple reports would contain acknowledgements of each other's existence, this does not always happen. Reviewers, therefore, have to be especially vigilant not to assume that independent studies have been reported.

Kagan (1992) included one very interesting Type 4 error involving double counting. This involved the case studies reported by Bullough and his colleagues. Bullough, Knowles, and Crow (1989) reported case studies of three beginning teachers named Bonnie, Lyle, and Helena. Then Bullough (1990) wrote about Helena alone, but called her Heidi, without mentioning that they were one and the same. Next, Bullough and Knowles (1990) wrote about Lyle alone, without mentioning the earlier write-up on Lyle. Finally, Bullough and Knowles (1991) wrote about Bonnie alone, except that this time Bonnie was called Barbara; it was revealed in an endnote that Barbara “chose to have her real name used” (p. 139). This endnote was not signaled at any place in the text where the name “Barbara” was used, but, rather, was added to another, quite unrelated note. Only by accident could a reader discover the information about the change of name. Kagan was apparently unaware of these strange occurrences and so treated these publications as though they were all reporting independent studies.

The effect of this Type 4 error was that Kagan (1992) claimed that six cases had been found, when there were in fact only three. To those who consider such matters as the replicability or frequency of occurrence of findings to be important, this error is significant.

Type 5: Nonrecognition of faulty author conclusions. Authors of original reports of research do not always represent their findings fully in their statements of conclusions. If reviewers uncritically accept such statements, they risk continuing the misrepresentation. The occurrence of this error can be contributed to by the original author, who may be biased to the extent of selectively incorporating only expected or hoped for findings in statements of conclusions.

One example of a Type 5 error that occurred in Kagan (1992) involved the report by Weinstein (1990). Weinstein studied all 38 student teachers enrolled in one section of an introductory course required for formal admission to a teacher education program in the northeastern United States. On the basis of responses to an initial questionnaire which asked “How well do you think you will do during student teaching?,” 15 were selected for interview. Students had also been asked to list the strengths and weaknesses they considered when answering that question. The 15 were then interviewed to explore topics in depth and answer questions about ways in which their thinking had changed during the semester. Of the 15, 12 agreed to be interviewed in the following semester. The questionnaire was administered in the autumn semester on the first and last days of class. Interviews were conducted in the following spring.

After reporting the results of her analyses of the data, Weinstein concluded, “What is most striking about the data reported here is the lack of change that
occurred during the semester” (p. 285). In Kagan (1992), Weinstein’s conclusion regarding change was repeated as follows: “Despite coursework and field experiences, the candidates’ beliefs about teaching and themselves as teachers remained unchanged throughout the semester” (p. 140). In fact, Weinstein’s findings were as follows:

(1) There was a significant decrease in optimism (p < .05), “although students remained extremely optimistic” (p. 282).

(2) “Explanations for the self-ratings given at the end of the semester were similar to those given at the beginning of the semester. Students continued to stress caring (32%) [down from 45%] and enthusiasm (40%) [up from 32%]. However, non-teaching experiences with children were rarely mentioned (40% at the beginning of the semester versus 2.6% at the end (p < .0001); instead, 18% of the students (in contrast to an initial 3%) now cited a knowledge of teaching gained from the course or the associated field experience (p = .03)” (p. 283).

(3) “Conceptions of a ‘really good teacher’ remained largely the same; however, there were some interesting changes. Ability to maintain discipline was now mentioned by 50% of the students (p = .05) [up from 37% ], as was enjoyment or enthusiasm for teaching (p < .01) [up from 23%]. Fewer subjects cited ability to motivate students (p = .09) [from 34% to 18%] and willingness to spend extra time and effort (p = .08) [from 32% to 16%]. There was also an increase in responses dealing with the ability to meet the diverse needs of individual students (from 5% to 19%, p = .07)” (pp. 284–285). Weinstein also found that there was a decrease in mentions of “professional behavior” from 10% to 0% (p < .05), but this was not mentioned in the text of her report.

None of these findings by Weinstein regarding change was mentioned in Kagan (1992), and so the conclusion was a gross misrepresentation of the facts, probably due to uncritical acceptance of Weinstein’s clearly unwarranted conclusion.

Type 6: Unwarranted attributions. This error consists of reviewers claiming that studies yield findings or reach conclusions that they do not. In an extreme form, this error can even consist of attributing findings to studies when the design limitations of the studies do not permit such findings to be reached, or when the studies do not even set out to investigate the subject of the attributed finding.

One glaring Type 6 error was found in Kagan’s (1992) treatment of the Hollingsworth (1989) report. In the review, it was said that Hollingsworth (1989) identified four factors that appeared to affect the acquisition of classroom knowledge by the novices: (a) their images of themselves as learners; (b) an awareness that they needed to temper initial beliefs and come to terms with classroom management; (c) the presence of a cooperating teacher who was a role model that facilitated growth; (d) placement with a cooperating teacher whose ideas and practices were somewhat different from the student teacher’s beliefs. Modeling seasoned teachers was not sufficient to promote conceptual change; cognitive dissonance was needed to force novices to confront and modify their personal beliefs. (p. 145)

This is what Hollingsworth (1989) wrote:

For those who did [reach a balanced managerial style], there seemed to be at least four explanatory factors that helped them acquire that knowledge: (a) a role image of themselves as learners and critics of teaching, which allowed

92
Errors in Synthesizing Research

for error and change; (b) an awareness that they needed to change their initial beliefs to come to terms with classroom organization; (c) the cooperating teacher and/or university supervisor as role models and facilitators of that change; and (d) a notion of having something worth teaching that demanded student cooperation. (p. 174)

Attributing to Hollingsworth (1989) the finding concerning cognitive dissonance was a Type 6 error in Kagan (1992).

Type 7: Suppression of contrary findings. Original reports sometimes contain findings that are actually contradictory of the generalizations which a reviewer claims they support. It is not necessarily the case in these instances that findings consistent with the reviewer's generalization are not contained in the reports; it is just that other, contradictory findings that are reported are ignored. Avoidance of Type 7 errors does not necessarily demand that reviewers acknowledge every single finding of a study, but it does require acknowledgement of every single finding that is contrary to a generalization a reviewer intends to make about the findings of the body of research concerned.

Grossman (1992) hinted that there was a Type 7 error in Kagan (1992) when she argued that in the part of her study (Grossman, 1989) “not directly addressed” by the review there was no evidence that preservice teachers complained that a particular course was “too theoretical” (p. 174).

Perhaps the most severe Type 7 error, however, was referred to above concerning the Gore and Zeichner (1991) study. Gore and Zeichner were concerned with levels of reflective thought exhibited by student teachers. Theirs was a case study concerned with action research and reflective teaching in preservice teacher education. The authors reported an analysis of the written reports of the action research projects conducted by 18 student teachers during the 1988–1989 academic year “to explore the extent to which action research seemed to be contributing to reflective teaching practice as we have defined it; that is, reflection within all three domains of rationality” (p. 129). The three domains were technical, practical, and critical rationality (Van Manen, 1977). Student teachers kept journals; conducted some formal observations of other classes; prepared, taught, and evaluated a unit of work; and conducted action research.

Gore and Zeichner (1991) discovered three broad groups of action research projects: (a) a small number in which there was “clear concern for moral and political issues as integral to the project” (p. 129), which they categorized as displaying critical rationality; (b) a larger number in which there was “some concern for these issues but [the writer] did not develop the ideas” (p. 129), which they categorized as displaying practical rationality; and (c) projects more than half of which “revealed no explicit concern for moral and political issues at all” (p. 129), which they categorized as displaying technical rationality.

In Kagan (1992) it was said that there was “little evidence of reflection; what little they did find consisted of technical rationality, the lowest level” (p. 142). This was not true. At least 6 of the 18 cases studied (Jo, Bruce, Melinda, Helen, Leslie, and Annette) were classified as displaying either critical or practical rationality, the two levels of reflection ranked above technical rationality. By denying the actual findings obtained and thereby committing a Type 7 error, the reviewer drew unwarranted conclusions about the influence of preservice teacher education upon student teachers’ cognitive performance.
The tertiary stage of a synthesis is that at which the reviewer seeks to assemble the evidence of the individual studies according to the main topics or issues investigated, in order to see whether meaningful and justifiable generalizations (syntheses) can be stated about them. The questions asked are, Do they add up?, and, If so, to what? Errors at this stage can lead to the statement of invalid generalizations and to the failure to recognize valid ones. Of course, errors made at the primary and secondary stages seriously threaten the validity of generalizations at the tertiary stage, but it is also possible that errors can emerge at this stage for the first time.

Type 8: Consequential errors. These are generalizations that are flawed as a consequence of errors made at earlier stages. Errors of any of the first seven types during the primary and secondary stages affect the validity of the findings of the body of studies that the reviewer attempts to synthesize at the tertiary stage. Colloquially, this problem has been captured in the well-known phrase “garbage in, garbage out!” Valid generalizations cannot be reached from erroneous conclusions about the individual studies that are part of a review.

The Kagan (1992) synthesis addressed the question “Do preservice candidates change their personal beliefs and images during the course of a teacher education program?” (p. 156) and constructed the generalization that “the personal beliefs and images that preservice candidates bring to programs of teacher education usually remain inflexible” (p. 154). Later, it was claimed that “all but one study indicated that personal beliefs remained stable” (p. 156). However, it has been demonstrated (Dunkin, 1995) that the reverse was the case and that most of the studies cited did find substantial change. It seems that the review’s representation of four of the studies (McDaniel, 1991; McLaughlin, 1991; Pigge & Marso, 1989; Weinstein, 1990) contained Type 7 errors, and that Type 6 errors had been made with another two (Calderhead & Robson, 1991; Gore & Zeichner, 1991), neither of which reported findings about change or lack of it. As a consequence of these errors, a Type 8 error was made in Kagan (1992) in the generalization about change in preservice teachers’ beliefs.

Type 9: Failure to marshall all evidence relevant to a generalization. When a reviewer, in the process of formulating conclusions, fails to recognize that a study contains evidence relevant to a generalization, he or she commits this type of error. This is different from a Type 1 error. It is not that an entire study is omitted from the review but that one or more of the study’s findings are not included in the assembling of evidence bearing upon the generalization in question.

In Kagan (1992), Type 9 errors were made when the reviewer failed to assemble evidence provided by Bullough et al. (1989), Grossman (1989), and Levin and Ammon (1992) concerning the generalization that university courses are not sufficiently relevant to the needs of student teachers and that such courses fail to provide novices with adequate procedural knowledge of classrooms. In the review, it was claimed that there was extensive support for that proposition (p. 162), but contrary evidence provided in the three studies cited above was ignored. The review failed to report Levin and Ammon’s finding that university courses were effective in securing growth in student teachers’ pedagogical conceptions. Furthermore, the review did not mention that Grossman had shown what could
happen in the absence of teacher education courses and had concluded that "teacher education coursework can help prospective teachers acquire knowledge about what students are likely to find difficult in a particular subject, and a realistic sense of students’ interests, abilities, understandings, and misconceptions concerning specific topics" (p. 206). Moreover, the review did not refer to the following conclusion reached by Bullough et al.:

All three teachers had adequate theoretical knowledge about teaching, exposure to and practice of appropriate teaching skills; indeed, they completed the same preservice teacher education program together. What they initially lacked were useful understandings of the contexts in which they would work and, particularly for Lyle, consistent, grounded, and accurate understandings of themselves as teachers. (p. 231)

Bullough et al. argued that teacher educators could do more to help prospective teachers answer the question “Who am I?” but that the understanding of school contexts was primarily the responsibility of school districts and principals.

The failure to bring the above findings by Bullough et al. (1989), Grossman (1989), and Levin and Ammon (1992) to bear on this generalization constituted a Type 9 error.3

Conclusions

The main concern to arise from this presentation of the types of errors facing reviewers is the trustworthiness of their syntheses. It has been argued that all nine types of error were present in just one synthesis (Kagan, 1992). If that is the case, it seems clear that no reliance should be placed on that synthesis. But does it say anything about other syntheses? Surely, readers of these works cannot go to the trouble of the detailed scrutiny required to check the validity of every synthesis before they decide whether or not to rely on them. Not even referees or editors can be expected to subject syntheses in manuscript form to the painstaking inquiry process required to establish their validity. In this respect, a synthesis of research is probably no different from the individual studies included in the review, all of which, themselves, are subject to error at every stage of their construction.

The only feasible, systematic approach to quality assurance in regard to syntheses of research is to educate educational researchers, all of whom conduct their own syntheses of research whenever they write a dissertation, a grant proposal, or a research report, to look for the types of errors identified here, and in the more general requirements of good scholarship. Moreover, potential users of syntheses should be encouraged to develop a healthy skepticism toward them. The availability of a typology of synthesizer errors should assist in both processes.

Notes

1It would be a breach of editorial ethics to disclose the identities of the authors of those papers.
2In Kagan (1992) it was mistakenly claimed that McLaughlin (1991) had been published in the Alberta Journal of Educational Research.
3A detailed analysis of errors found in Kagan (1992) concerning research on first-year and beginning teachers is to be found in Dunkin (in press).
References


**Author**

MICHAEL J. DUNKIN is Professor, School of Teacher Education, The University of New South Wales–St. George Campus, P.O. Box 88, Oatley, NSW, Australia 2223; m.dunkin@unsw.edu.au. He specializes in research on teaching and teacher education.

Received March 7, 1995
Revision received January 16, 1996
Accepted March 19, 1996