Afterword: Key Questions for Thought and Action

Lorin W. Anderson
University of South Carolina
United States

Abstract: In this summary article, six recommendations for the design, implementation, and interpretation of educational evaluations are presented and discussed. These recommendations are based on common "threads" that run through most, if not all, of the papers included in this special issue. The recommendations concern (1) the need for awareness of the political, societal, cultural, and economic factors affecting evaluation studies; (2) the importance of knowing and understanding the stakeholder groups; (3) the need to ensure that the purposes of the evaluation are explicit and clear; (4) the importance of allowing flexibility in the implementation of the evaluation when needed to account for issues that arise during the evaluation process; (5) the need to ensure that the data gathered as part of the evaluation process are of the highest technical quality possible; and the (6) importance of ensuring that the results of the evaluation are interpreted correctly and well understood by stakeholders and decision makers.

Keywords: Context; stakeholders; evaluation; mixed methods; technical quality; interpretation; outcomes

Cuestiones clave para el pensamiento y la acción
Resumen: En este capítulo final se ofrecen y discuten seis recomendaciones para diseñar, implementar e interpretar evaluaciones educativas. Estas recomendaciones se basan en las líneas comunes que corren a lo largo de la mayoría, si no es que de todos, los artículos incluidos en este número especial. 1) La necesidad de tener conciencia de los factores políticos, sociales, culturales y económicos que afectan a los estudios de evaluación: 2) la importancia de conocer y hacer explícitos a los actores involucrados; 3) la necesidad de asegurar que los propósitos de la evaluación sean explícitos y claros; 4) la importancia de permitir flexibilidad en la implementación de la evaluación cuando sea necesario dar cuenta de los tópicos de discusión que surjan durante el proceso de evaluación; 6) la Importancia de asegurar que los resultados de la evaluación sean interpretados correctamente y sean debidamente entendidos por los actores participantes y los tomadores de decisiones.
Palabras-clave: Contexto; evaluación; métodos mixtos; calidad técnica; interpretación; resultados

Questões-chave para pensamento e ação
Resumo: Este capítulo final oferece e discute seis recomendações para projetar, implementar e interpretar avaliações educacionais. Estas recomendações são baseadas em linhas comuns que percorrem a maioria, se não todos, os artigos desta edição especial. 1) A necessidade de estar ciente dos fatores políticos, sociais, culturais e económicos que afectam os estudos de avaliação: 2) a importância de conhecer e explicitar os atores envolvidos; 3) a necessidade de garantir que os objetivos da avaliação sejam explícitos e claros; 4) a importância de permitir flexibilidade na implementação da avaliação, quando necessário para dar conta dos tópicos de discussão que surgem durante o processo de avaliação; 6) a importância de assegurar que os resultados da avaliação são interpretados corretamente e estão devidamente compreendido pelos atores participantes e decisores.
Palavras-chave: Contexto; avaliação; métodos mistos; qualidade técnica; interpretação; resultados

Afterword: Key Questions for Thought and Action

The authors included in this special issue represent four countries: Canada, Mexico, South Africa, and the United States. The articles differ in their focus (educational systems, programs, teachers, and students). Some contributions are primarily conceptual, others are empirical, and still others are a little of each. Despite these differences, however, several threads run through them. In this closing article, these "threads" are represented as a set of six recommendations for planning, executing, and interpreting sound, defensible evaluations.

The rich array of theoretical and methodological perspectives offered by the authors is intended to help those preparing new evaluation proposals place those proposals in a larger framework and to understand the factors that influence the design, implementation, and success of evaluations. Many of the authors (Anderson, Berliner, Philips, Schubert, and Shavelson) provide historical overviews intended to enable educators, evaluators, and policy makers to learn from history, rather than repeat its mistakes.

Understand Political, Societal, Cultural, and Economic Influences

> *Evaluations are themselves interventions—sometimes large-scale and expensive ones—and they often, if not always, take place in a social context that is marked by discord* (Phillips, this issue)

At the heart of the noun *evaluation* is the root word *value*. Schubert's article reminds us that "all choices one makes, every action engaged, involve values; thus, they pertain to evaluation. Even if the values are not addressed, they govern by default, chance, or expediency."

In most countries throughout the world values pertaining to education are not left to chance or expediency. Rather, laws are passed and regulations enacted that, in essence, make these values explicit. What are the primary aims of education? What should students learn? How should schools be organized? What materials should be included in (or excluded from) the intended curriculum? What qualifications and training do teachers need? How should students be assessed and evaluated? Since all of these are, in fact, questions of value, it should not be surprising that not everyone arrives at the same answers.

This is where politics enters the picture. Politics is defined in the dictionary as the "activities associated with the governance of a country or other area, <u>especially the debate or conflict among individuals or parties having or hoping to achieve power</u>." In the field of education, political factors are involved in the design, selection, and funding of policies, projects, plans, programs, and evaluations. As Shavelson's contribution notes, "Politics matter a great deal. Ignore politics at your peril." To complicate matters further, answers given to these questions are likely to change, often dramatically, as elected or appointed officials (or entire administrations) are replaced. De Ibarrola's contribution is an excellent, comprehensive case study of the effect that conflicting values have had on educational reform and teacher evaluation in Mexico.

Turning to social factors, Ercikan, Asil, & Grover's article reminds us of the importance of the so-called "digital divide." They define digital divide as "social inequity between individuals regarding (a) access to information and communication technology (ICT), (b) frequency of use of technology, and (c) ability to use ICT for different purposes." This divide has become increasingly important as states, regions, and nations move to conducting student assessments electronically. Other important social factors mentioned by chapter authors are gender and socio-economic status (SES). As Van der Berg's article on international comparisons makes quite clear, there are large SES differences *within* countries and regions of countries and *across* countries and regions.

Borrowing from the fields of anthropology and sociology, a culture consists the beliefs, values, behaviors, traditions, and customs that are shared among, and accepted by, people in a particular society and which are transmitted from one generation to the next. In virtually every country, there is a dominant culture. The dominant culture can often be ascertained by perusing secondary school history textbooks. These textbooks contain what Schubert refers to as the "master narrative." Within the dominant culture, virtually all countries have subcultures, that is, the behaviors and beliefs characteristic of a <u>particular social, ethnic, or age group</u>. Schubert describes numerous subcultures that have been marginalized and often oppressed within many countries (e.g., indigenous populations, women, racial minorities).

Schubert suggests that the presence of these subcultures "challenge the field of education with questions about diversity, relative to race, class, gender, ethnicity, sexuality, (dis)ability, language, culture, tradition, place, and more. Whose voices should be heard? Whose values should be a basis for curriculum and evaluation? Who benefits from and who is harmed by past and present practices?" In response to these questions, Schmelkes describes the efforts made by Mexico's National Institute for the Evaluation of Education (INEE) to create an intercultural approach to educational evaluation.

Although not explicitly addressed by any of these authors, economic factors are implicit in several. Economic factors include a variety of resources (e.g., financial, human, material) which can be made available as needed. Phillips recounts a story of Lee Cronbach asking members of an evaluation team whether after the expenditure of vast amounts of money it was likely that the program would be shut down. Their answer, "No!" That is, regardless of the results of the evaluation, the program would quite likely continue. As an interesting contrast, Shavelson points out that despite evaluation results that supported the positive benefits of smaller classes for young children, the state legislature decided <u>not</u> to reduce class size statewide because of the cost involved. Berliner argues that one of the main reasons for including student achievement test scores as part of teacher evaluation is that doing so is inexpensive since the scores are available and easily accessible. Although one of the key elements of teacher evaluation reform in Mexico was a Technical Assistance Service System near to each school, the system has not yet been put in place, largely because of the tremendous cost involved and the lack of the human resources.

When considering educational evaluations, then, one would be wise to ask the following questions:

1. Who are the power brokers? What and where are the potential political conflicts?
2. What social and cultural factors should we take into consideration as we plan the evaluation?
3. What are the key components of the evaluation plan, and how and when should each component be implemented?
4. What is the scale of the intended impact of the evaluation (national, regional, local)? If multiple levels are involved, can the data be collected and analyzed in ways that disentangle the various levels?
5. What are the cost factors of both the evaluation itself and the resources needed to follow up on the evaluation? Are there sufficient resources available to produce a high quality, useful evaluation? In the case of formative evaluation (see below), are sufficient financial and human resources available to make and sustain the changes suggested by the evaluation results?

Understand the Leadership and Influence of Stakeholder Groups

> *Over the decades many evaluation reports have been ignored or pushed aside, because they focused on an issue or a function that was not the main concern of the stakeholders—that is, they gathered information that was not relevant to the real decisions that the stakeholders were interested in making* (Phillips, this issue)

The term "stakeholder" typically refers to an individual or group that is invested in the welfare and success of an organization, particularly the programs and personnel within that organization. In education, stakeholders have a "stake" in the educational system because of personal, professional, civic, or financial interests or concerns. These stakeholders may include administrators, teachers, students, parents, policy makers (including school board members and elected or appointed officials of different ranks), community leaders or agencies, members of advocacy groups, technical advisors, and members of the media. In some situations, there are likely to be groups that are influenced greatly by evaluative decisions (often negatively), but who are not considered as stakeholders, either by themselves or by others. If possible, evaluators should seek out these groups to ensure that <u>all</u> relevant stakeholder groups are included in the evaluative process. [In her article, de Ibarrola uses "actors," rather than "stakeholders," but for practical purposes the concepts are quite similar.]

The relevance of particular stakeholder groups depends to a great extent on what and who are being evaluated. For example, parents are very important stakeholders when students are being evaluated (Anderson), but are likely to be less important when teachers are being evaluated (Berliner). Similarly, public officials are more likely to be important stakeholders when funding is needed for a new program, but less likely to be important when modifications requiring little, if any, additional cost are made to existing programs. Schmelkes provides an excellent example of the way in which intercultural teachers can become important stakeholders in planning and implementing their own evaluation.

The reader will note that this second recommendation contains two verbs: know and understand. These verbs can be translated into two questions: Who are the stakeholder groups (know)? What are their interests and concerns (understand)? Both questions should be answered as early in the evaluation process as possible. Once these questions have been answered, representatives of appropriate stakeholder groups should be involved in the evaluation from the outset. De Ibarrola's contribution is an object lesson on the failure to do so. She hints at the possibility that misunderstanding among stakeholders is not only a matter of vested interested interests, but may have deep historical antecedents. Schubert) reminds us that different stakeholder groups may hold different curricular orientations, some of which are implicit and therefore difficult to discern.

Finally, stakeholders, although critically important to the success of an evaluation, are not necessarily the people making the evaluative decisions. The issue of who <u>should</u> make the decisions has been debated for at least a half century and is described in great detail by Phillips. Some argue that those responsible for the design, conduct, and oversight of the evaluation should be the ultimate decision makers. The rationale is that they know more about the "ins" and "outs" of the evaluation process and outcomes than anyone else. Others argue that the decisions should be made by those who have the <u>authority</u> to make decisions within the organizational context within which the evaluation takes place. That is, those who plan and carry out the evaluation are expected or encouraged to make recommendations, but ultimately the decisions are made by people in positions of authority.

Increasingly, the responsibility for designing, conducting, and providing oversight for evaluations is given to a group of people within an institution (e.g., a university) or an agency (e.g., INEE). When this is the case, it makes more sense for this group to serve in an advisory capacity, rather than as decision makers.

The decision maker (or decision makers) should be made known to everyone. Once identified, the relationship between decision makers(s) and those responsible for conducting and executing the evaluation as well as with representatives of relevant stakeholder groups should be clarified and made explicit. To facilitate regular interactive communication, advisory groups that meet on a regular basis throughout the evaluation process are often quite useful.

Finally, building consensus among stakeholder groups must take into consideration the consequences of the evaluation. Do the results of the evaluation affect the future of students (Anderson), the continued employment of teachers (Berliner), or needed modifications of legislative or educational programs (de Ibarrola, Shavelson)? A shared understanding of the consequences early in the process is likely to lead to a more productive, less contentious discussion of the results.

Ensure the Purposes of the Evaluation Are Explicit

> *Evaluations [of teachers] can help in the design of staff development, can inform teacher training institutions about some deficits they have, and … assist personnel decision by principals, personnel officers, superintendents, or school boards* (Berliner, this issue).

Historically, there are two general purposes (or functions) of evaluation: summative and formative. Although much has been written on each of them and their differences, the example provided by Bob Stake (as cited by Phillips) captures the distinction quite well. "When the cook tastes the soup, this is formative evaluation; when the customer tastes the soup, this is summative evaluation." In Berliner's quote above, using teacher evaluations to design staff development and inform teacher-training institutions of their deficits are examples of formative evaluation. Using evaluations to assist in making personnel decisions (e.g., continued employment, salary increases, awarding of tenure) is an example of summative evaluation. As a set the chapters provide insights into the many functions that evaluation can serve: evaluation of educational outcomes consistent with various curricular orientations (Schubert), evaluation of students to motivate them or to improve the instruction they receive (Anderson), evaluation of educational programs to describe what is happening or to make causal inferences (Shavelson), evaluation of student achievement for the purpose of improving national systems of education (Van Der Berg).

There is a consensus among the authors that the emphasis should be on formative evaluation. Formative evaluation is more likely to offer the possibilities of identifying "unintended effects" (that is, what actually happened rather than what should have happened) (Phillips). While the focus of summative evaluation is almost exclusively on the intended curriculum (that is, the curriculum as described in courses of study or lists of academic standards), formative evaluation enables the evaluator to gather information about the hidden curriculum, experienced curriculum, outside curriculum, and null curriculum (Schubert). Shavelson argues that formative evaluation is needed to ensure a program is in "consistent working order" before examining program (causal) effects (summative evaluation). Finally, Schubert suggests that formative evaluation, when done well, can "further the educative process" by redirecting emphasis before too much damage can be done by continuing the implementation of a program that seems valuable in the abstract, but is found to be ineffective or harmful in practice.

Regardless of the stated purpose(s) of evaluation, one should be aware of the motive underlying the evaluation. "Motive" as defined in the dictionary is a "reason for doing something, especially one that is hidden or not obvious." Consider, for example, a teacher evaluation program that is intended to assist in the making of personnel decisions (see above). Within this general purpose, the focus may be on identifying and rewarding the best teachers or identifying and eliminating the worst teachers. Berliner suggests that the motive underlying much of teacher evaluation in the United States is to "get rid of 'bad' teachers." Within the Mexican context, although one of the stated purposes of the reform efforts was the "professionalization of teachers and officials," a primary motive of the government was to "recover [federal] control of national education" from the hands of the National Teachers' Union (De Ibarrola, Shavelson).

One of the potential drawbacks of formative evaluation is that the costs involved in making the improvements that are needed based on the results of the evaluation are often quite high. For example, a study conducted by the New Teacher Project in the United States in 2015 estimated that the cost of teacher professional development, a strategy often recommended for teachers found to be deficient based on the results of their evaluations, was approximately $18,000 per teacher. At the same time, however, simply pointing out problems or inadequacies without attempting to remedy them decreases the utility of evaluation, making it more of an exercise than a meaningful activity. Without sufficient funds being allocated, then, it should not be surprisingly that few, if any, of the legal requirements concerning specific training programs and the technical assistance service system in Mexico were implemented successfully, if they were implemented at all.

Understand Who and What will be Evaluated

> *As Bertrand Russell noted long ago, striving to create good education is intricately connected to the quest for the good life. Accepting the importance of the notion of good education and education for the good life as basic is only a beginning. Alternative meanings of what good education and good life mean must also be addressed by evaluators* (Schubert, this issue).

If we are evaluating a program that purports to improve reading achievement, do we get information only from students (e.g., test scores) or do we include others in our evaluation design (e.g., reading teachers, teachers in other subject areas, parents)? If we are evaluating teachers, do we evaluate all teachers, only novice teachers, only teachers of academic subjects, only teachers of exceptional children? Also, is teacher participation mandatory or voluntary? (de Ibarrola). If we are evaluating the relative effectiveness of school systems in various countries, do we include all students, only students of a particular age or enrolled in a particular grade, only students who were tested? (See Van der Berg's article). The first order of business, then, is to decide specifically who will be included in (and excluded from) the evaluation and how that decision is likely to limit or expand the recommendations that might be made at the end of the study.

If we are going to evaluate a program designed to improve reading achievement, do we gather information on oral reading fluency, vocabulary, literal comprehension, making inferences from reading material, interest in reading, or confidence as a reader? If we are going to evaluate teachers, do we include measures of teacher competence, performance, or effectiveness? Competence is what a teacher knows and can do, performance is what a teacher actually does, and effectiveness is the impact that knowing and doing has on students. If we are going to evaluate the relative effectiveness of school systems in various countries, do we include on the assessment test items focusing on recall of facts, application of skills, or the ability to solve problems? If the application of skills, do we sample from all skills or from those skills common to the curricula of all countries? All of these questions pertain to the "what" of evaluation.

Why is it important to clearly specify the "what" of evaluation? Consider the data summarized by Berliner. He begins by describing the complex act of teaching and the lack of sensitivity of many of the instruments used in teacher evaluation. He then reports a weak correlation between value-added measures (those based on perceived increases in students' test scores) and observational data (based on classroom visits). If these are two measures of the same teaching variable(s), then such a low correlation would be unexpected. On the other hand, it may be that two different teaching variables are being measured. Value-added measures are purported to be measures of teacher effectiveness, whereas observational data are purported to be measures of teacher performance. If this is the case, the low correlation suggests that teachers who "perform" at higher levels are not necessarily more effective than those who perform at lower levels.

Before moving on to the next recommendation, a comment on the clause in the recommendation that appears after the comma is noteworthy. Evaluations take time, with data often collected at various points during that time interval. Based on evidence gathered during that time interval, guiding questions may need to be modified and/or evaluation procedures (including data collection instruments) may need to be altered (see de Ibarrola). As a specific example, suppose that early in an evaluation it becomes evident that the interview protocol prepared for parents and community leaders is asking the wrong questions or asking the questions in the wrong way. Should the interview protocol be modified to increase the quality of the responses? In the early years of program evaluation the answer would likely be "No." One of the hallmarks of the quality of these early evaluations was something called "fidelity of implementation." That is, to what extent is the program implemented as it was designed. With an emphasis on "fidelity of implementation," it was

generally unwise to vary a great deal from the original evaluation plan. Increasingly, however, such modifications have been seen as not only useful, but necessarily, to ensure the highest quality data possible. The downside is that such changes make it difficult if not impossible to compare data collected at one point in time with data collected at another. At the same time, however, of what benefit is there in comparing data of questionable validity or utility? This question leads us nicely to the fifth recommendation.

Ensure that High Quality Data Are Collected

> *What is measured and how it is measured impacts, in significant part, what is found; change the measurement and findings may change … [At the same time, however] evaluation methods should not drive the evaluation. Rather the questions that gave rise to the evaluation should drive the design and conduct of the evaluation* (Shavelson, this issue)

The authors offer several suggestions for increasing the likelihood that the data (also referred to as information or evidence) collected during the evaluation are of sufficient technical quality. That is, the data should be sufficiently valid, reliable, and useful.

First, Shavelson suggests that decisions about data collection instruments should be made <u>after</u> the questions have been established and agreed upon. "Reliability, validity and utility must be aligned [with] the measurement's intended purpose." Far too often (as in the case of using student test scores in teacher evaluation) instruments are chosen simply because they are readily available. Beginning with the questions derived from the evaluation as conceptualized and designed makes it more likely that appropriate and relevant instruments and measures will be selected or developed.

Second, Shavelson suggests that mixed data collection methods should be used whenever possible. Examples of available methods include observations (both participant and structured), open-ended interviews, structured surveys, document analysis, and/or standardized tests. When mixed methods are used, it is possible to examine the relationships among the data collected to determine the extent to which the data are consistent across methods. Consistency, when it exists, gives the evaluator greater confidence in the data. When inconsistency is evident (as in the case of Berliner's teacher evaluation data), evaluators must search for explanations for the inconsistency. Either way, the use of mixed methods permits a more complete understanding of the phenomena being investigated; this increased understanding, in turn, is likely to result in more informed, defensible judgments.

Third, it is a good practice to have data collection instruments and methods reviewed prior to using them in the evaluation. The review may include a relatively inexpensive field test and/or convening a panel of judges who examine the instruments and methods in some detail. If a panel is used, it should include representatives of different stakeholder groups. If student test scores are to be used to evaluate teachers, then the instructional sensitivity of the items (that is, the extent to which the items are indeed sensitive to variations in instructional quality) should be examined (Berliner). Having data collection instruments and methods criticized <u>after the evaluation has been completed</u> raises questions about the credibility of the results of the entire evaluation and quite likely requires a careful review of the entire evaluation process (de Ibarrola).

Fourth, the unit of analysis or unit of aggregation can, and often does, impact on estimates of technical quality. As an example, Anderson finds that although the reliability of grades assigned to students on a single assignment is quite low, the reliability of grades assigned to students based on a cumulative set of assignments is reasonably high. Similarly, Berliner reports that although one or two observations of teachers is unlikely to produce reliable data, data from eight or more observations can produce a level of reliability that is sufficient to support sound, reasonable evaluations.

Finally, Van der Berg suggests that the major criticisms of evaluation studies pertain to the data collection methods. Specifically, critics focus on:

1. the validity and reliability of standardized testing,
2. reliance on quantitative measures, and
3. an emphasis on measurable aspects of education only.

Ensure Proper Interpretation of Evaluation Results

> *The existence of a digital divide … points to a possible widening of achievement gaps on assessments which may not be true reflections of group differences in knowledge, skills and competencies*
> (Ercikan, Asil, & Grover, this issue)

Stated somewhat differently, Ercikan and her colleagues are suggesting that if there are differences among students in terms of their access, frequency of use, and ability to use information and communication technology (ICT), and if standardized achievement tests are administered using this technology, the interpretation that students differ in their actual or "true" level of tested achievement may not be valid.

In evaluation, more valid interpretations are most often made when data are disaggregated so that comparisons among relevant groups and subgroups can be examined. In his article, Van der Berg illustrates this point quite well with an example from the 2012 PISA test data. To be administered the test, students had to be 15 years old and be at least in grade 7. It is worth noting that the differences among participating countries in the percentage of 15-year-old students who have not yet reached grade 7 is quite large (as high as 50%). As Van der Berg points out, however, the "implicit working assumption … is that those 15 year olds who have not reached at least grade 7 … have not achieved basic numeracy." This assumption impacts greatly on the interpretation of student scores in individual countries. For example, when 15-year-olds not reaching grade 7 are assigned to the "below basic numeracy" category, the performance of students in Viet Nam who achieved "basic literacy" ranked well below the international average. However, more than 40% of Viet Nam's 15-year old students had not yet reached grade 7. If these students are excluded from the analysis rather than being included in the "below basic numeracy" group," the performance of Viet Nam students ranks in the upper quarter of the distribution of countries. Based on this analysis (as well as other analyses in his article), Van der Berg concludes, "evaluations contribute more to our understanding of educational deficits in developing countries when they are combined with data on access and coverage." Shavelson and Ercikan, Asil, & Grover argue for the importance of disaggregating student data by variables such as prior achievement, socioeconomic status (SES), and gender. Similarly, disaggregating teacher data by experience, education level, and gender may lead to more appropriate interpretations of the data.

The second part of this recommendation emphasizes the importance of making sure that stakeholders and decision-makers understand the data as interpreted. Anderson suggests that educators, particularly those involved in collecting and disseminating the results of evaluations (specifically student grades) must "find ways to communicate … so that the information needs of a variety of audiences are met." He then adds, "Rather than assume they understand the information needs of various audiences, educators would be wise to ask them." Based on the collective experience of the chapter authors, what and how evaluation results are reported to parents, elected officials, and members of the media must be quite different if they are to be communicated effectively. Parents are more concerned about their children or their school, not children or schools in general. Elected officials want information to be presented as briefly as possible and displayed as bullet points. The media prefers information that provides useable quotations. How to meet these

needs while at the same time being true to reasonable and defensible interpretations of the data is a challenge facing virtually every evaluator.

Concluding Statement

The contributions in this special issue are intended to help those interested or involved in educational evaluation learn from our collective successes and failures. The recommendations included here are best seen as guideposts to use in the design, execution, and interpretation of education evaluations.

Thomas Alva Edison said that genius is 1% inspiration and 99% perspiration. A similar statement can be said about education evaluation (perhaps with slightly different percentages). Good design often requires inspiration—insight, creativity, and cleverness. Execution, however, is a painstaking process that requires attention to detail, finding knowledgeable respondents, catching and correcting mistakes, entering and analyzing data, ensuring the proper interpretation of results, and working to ensure that the results are communicated in ways that facilitate understanding and, where possible, improved practice.

Because evaluation is time-consuming, exacting, and can often be tedious, one might ask whether the enterprise of evaluation is worth the time, effort, and expenditure of funds. To address this wondering, we quote Van der Berg (this issue): "The efficient and targeted application of resources and of policies cannot take place in an information vacuum; they require information on system performance, inequalities, progress, and stagnation that can only be gleaned from wide ranging data gathering and interpretation processes."

About the Author

Lorin W. Anderson
University of South Carolina (Emeritus)
anderson.lorinw@gmail.com
Lorin W. Anderson is a Carolina Distinguished Professor Emeritus at the University of South Carolina, where he served on the faculty from August, 1973, until his retirement in August, 2006. During his tenure at the University he taught graduate courses in research design, classroom assessment, curriculum studies, and teacher effectiveness. He received his Ph.D. in Measurement, Evaluation, and Statistical Analysis from the University of Chicago, where he was a student of Benjamin S. Bloom. He holds a master's degree from the University of Minnesota and a bachelor's degree from Macalester College. Professor Anderson has authored and/or edited 18 books and has had 40 journal articles published. His most recognized and impactful works are *Increasing Teacher Effectiveness, Second Edition*, published by UNESCO in 2004, and *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, published by Pearson in 2001. He is a co-founder of the Center of Excellence for Preparing Teachers of Children of Poverty, which is celebrating its 14th anniversary this year. In addition, he has established a scholarship program for first-generation college students who plan to become teachers.

About the Guest Editors

Lorin W. Anderson
University of South Carolina (Emeritus)
anderson.lorinw@gmail.com
Lorin W. Anderson is a Carolina Distinguished Professor Emeritus at the University of South Carolina, where he served on the faculty from August, 1973, until his retirement in August, 2006. During his tenure at the University he taught graduate courses in research design, classroom assessment, curriculum studies, and teacher effectiveness. He received his Ph.D. in Measurement, Evaluation, and Statistical Analysis from the University of Chicago, where he was a student of Benjamin S. Bloom. He holds a master's degree from the University of Minnesota and a bachelor's degree from Macalester College. Professor Anderson has authored and/or edited 18 books and has had 40 journal articles published. His most recognized and impactful works are *Increasing Teacher Effectiveness, Second Edition*, published by UNESCO in 2004, and *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, published by Pearson in 2001. He is a co-founder of the Center of Excellence for Preparing Teachers of Children of Poverty, which is celebrating its 14th anniversary this year. In addition, he has established a scholarship program for first-generation college students who plan to become teachers.

Maria de Ibarrola
Center for Research and Advanced Studies
mdeibarrola@gmail.com
Maria de Ibarrola is a Professor and high-ranking National Researcher in Mexico, where since 1977 she has been a faculty-member in the Department of Educational Research at the Center for Research and Advanced Studies. Her undergraduate training was in sociology at the National Autonomous University of Mexico, and she also holds a master's degree in sociology from the University of Montreal (Canada) and a doctorate from the Center for Research and Advanced Studies in Mexico. At the Center she leads a research program in the politics, institutions and actors that shape the relations between education and work; and with the agreement of her Center and the National Union of Educational Workers, for the years 1989-1998 she served as General Director of the Union's Foundation for the improvement of teachers' culture and training. Maria has served as President of the Mexican Council of Educational Research, and as an adviser to UNESCO and various regional and national bodies. She has published more than 50 research papers, 35 book chapters, and 20 books; and she is a Past-President of the International Academy of Education.

D. C. Phillips
Stanford University
d.c.phillips@gmail.com
D. C. Phillips was born, educated, and began his professional life in Australia; he holds a B.Sc., B.Ed., M. Ed., and Ph.D. from the University of Melbourne. After teaching in high schools and at Monash University, he moved to Stanford University in the USA in 1974, where for a period he served as Associate Dean and later as Interim Dean of the School of Education, and where he is currently Professor Emeritus of Education and Philosophy. He is a philosopher of education and of social science, and has taught courses and also has published widely on the philosophers of science Popper, Kuhn and Lakatos; on philosophical issues in educational research and in program evaluation; on John Dewey and William James; and on social and psychological constructivism. For several years at Stanford he directed the Evaluation Training Program, and he also chaired a national

Task Force representing eleven prominent Schools of Education that had received Spencer Foundation grants to make innovations to their doctoral-level research training programs. He is a Fellow of the IAE, and a member of the U.S. National Academy of Education, and has been a Fellow at the Center for Advanced Study in the Behavioral Sciences. Among his most recent publications are the *Encyclopedia of Educational Theory and Philosophy* (Sage; editor) and *A Companion to John Dewey's "Democracy and Education"* (University of Chicago Press).

education policy analysis archives

Volume 26 Number 55   April 16, 2018   ISSN 1068-2341

Please send errata notes to Audrey Amrein-Beardsley at Audrey.beardsley@asu.edu

Join EPAA's Facebook community at https://www.facebook.com/EPAAAAPE and **Twitter feed** @epaa_aape.