

Aggregating school based findings to support decision making: Implications for educational leadership

Theodore S. Kaniuka

Fayetteville (NC) State University

Michael R. Vitale

East Carolina University

Nancy R. Romance

Florida Atlantic University

Successful school reform is dependent on the quality of decisions made by educational leaders. In such decision making, educational leaders are charged with using sound research findings as the basis for choosing school reform initiatives. As part of the debate regarding the usability of various evaluative research designs in providing information in support of decision making, randomised field trials (RFT) have been advanced as the only valid way of determining program effectiveness. This paper presents a methodological rationale that would apply multi-level statistical analysis to aggregated, pre-post intervention data readily available from multiple school sites within a multiple baseline design framework to provide educational leaders with a valid alternative to RFT designs. Presented and discussed is an illustrative application of the potential value of such a design, in establishing the effectiveness of a cluster of reading programs in a form, that is directly applicable by educational decision makers considering reform initiatives and involving developmental reading.

Introduction

Educational leaders have been tasked with using evidence-based interventions to improve student learning as predicated upon school reforms such as *No Child Left Behind* (2002). While seemingly straightforward, the reality of using evidenced-based interventions is highly problematic for a number of reasons (Cooper, Levin & Campbell, 2009; *Education Week*, 2013; Nutley et al., 2007; Rickinson, 2005). In addition to barriers that exist at the practitioner level, researchers often fail to agree what constitutes valid research and how research findings can be used to support school reform (Bulterman-Bos, 2008; Chatterji, 2008; Slavin, 2008a; Sloan, 2008). At the same time, recent refinements to alternatives to the randomised field trail (RFT) design have been used to demonstrate the plausibility of using quasi-experimental designs as efficient and inferentially-effective alternatives across disciplines. (Lesik, 2006; Moss & Yeaton, 2006; van der Heyden, Witt & Gilbertson, 2007; West et.al. 2008).

This paper argues that the use of research to support educational decision making can be enhanced by applying the logical framework of multiple-baseline designs through the large-scale multilevel statistical analysis (Raudenbush & Bryk, 2001) of school-based evaluative data that report pre- and post intervention achievement findings. By applying multiple-baseline design logic to such multi-site data, the finding of multilevel statistical analyses of such data could provide the type of information about innovation effectiveness that, ultimately, is of greater direct utility to educational leaders than RFT

experimental studies which, because of their cost, are highly limited. Therefore, the question raised in this paper is about constructing an informational framework that educational decision makers can use as a valid alternative to RFT. And, by implication, if such a valid informational framework exists, are RFT studies necessary (vs. just sufficient) for making sound instructional decisions?

School reform decisions and the use of research

An important methodological emphasis in recent years has been the systemic adoption of randomised field trials (RFT) as a standard for evaluating the effectiveness of instructional interventions in schools (e.g. standards per IES *Requests for applications*, 2013) and the associated multi-level statistical modeling of the data obtained in such designs (e.g., Raudenbush & Bryk, 2001). In general, there is little argument regarding the soundness of the foundations for either RFT designs or for multi-level statistical modeling per se. Because school reform is necessarily an evolutionary process, the emphasis on RFT as the primary means for drawing evaluative conclusions regarding the effectiveness of instructional interventions artificially limits the decision-making of school leaders about the potential value of different instructional interventions (e.g., Burch, 2007; Slavin, 2008a, 2008b; Sloane, 2008; Stuart, 2007). However, for cases in which replicated studies using RFT for a specific intervention exist, sound estimates of the effects certainly can be obtained. But, at the present time, the availability of such findings is limited and, by inference, excludes from consideration a wide range of interventions whose outcomes have the potential for engendering the systemic improvement of educational outcomes. By emphasising the research standard of replicability, this paper argues that descriptive achievement data fitting a multiple baseline design framework can provide causal conclusions about the instructional effectiveness of specific interventions in a form useable by instructional decision makers.

The use of research by educational leadership to support policy decisions and school reform has a problematic history (Lagemann, 2002) that has resulted in the educational profession being accused of not being a research-based profession (see Hess, 2007). Recently Levin (2010) suggested that to increase the use of research by school leadership, focusing on the social aspect of how leaders access and evaluate information is an important element in the decision making process. The social aspect of how people access information has been studied in depth. For example, in summarising research on the diffusion of innovations, Rogers (2003) commented that people often rely on non-scientific sources as they make decisions about adopting innovations. Other researchers supporting this view found that managers and other leaders rely on their own experiences and opinions of colleagues more than on research evidence (Dobbins et al. 2007; Maynard 2007). When making decisions, educational leaders behave similarly in seeking the advice of colleagues, reflect on their own experiences, and depend on localised knowledge (Kochanek & Clifford, 2011). Contributing to the reluctance of many practitioners to use research and scientific knowledge is also that they often find the style of research presentations difficult to understand, regarding the way it is conducted as de-contextualised, and its relevance is severely limited (Fusarelli, 2008). The question, then, is

what can researchers do to improve the access, understanding, and ultimate use of research findings by educational leaders?

Considering the above, Coburn, Honig and Stein, (2008) offered a comprehensive review of the literature on district level use of research and evidence during decision-making. While they argued that researchers needed to continue to provide high quality research, they also noted that researchers also needed to consider district variables that often determine if and how research evidence is used. Specifically, in support of this view, Coburn et al. suggested that researchers needed to adopt an important role of “supporting the development of district capacity to effectively engage in research activities so that district personnel attend to and access this research.” p. 29.

While central to the purpose of educational research, the idea of developing causal arguments can only yield potential value if they motivate people to action. William and Lester (2008) offered the view that to actualise research into action, moving away from the generation of “knowledge” and theories toward inspiring individuals to action is fundamental. The problem for educational researchers, then, is to present the information they produce in a manner that inspires educational practitioners into action. This idea was discussed by Flyvbjerg (2001) who argued that research needs to be considered a phronetic activity, that is, an activity that induces action. He stated:

... phronetic social science is problem-driven and not methodology-driven, in the sense that it employs those methods which for a given problematic best help answer the four-value rational questions [Where are we going? Who gains and who loses, and by which mechanisms of power? Is this development desirable? What, if anything, should we do about it?]. (p. 196)

Clearly, one interpretation of the overly-strong emphasis on RFT is that of methodology over problem-solving that calls into question the relevance of adhering to this methodology alone when alternatives exist that also offer valid answers which motivate people to action. Furthermore, the engendering of actions through the dynamics of knowledge must necessarily incorporate the circumstances in which it is developed to be transferable across educational contexts to meet the localised requirement many educational leaders need in order to see research as relevant.

Beyond the relevancy concern is one of professional ethics. For example, in addition to providing an approach for documenting the effectiveness of educational interventions by using aggregated school-based evaluation, the methodological approach presented in this paper also provides a localised context for interpretation of the findings by practitioners. In doing so, positive achievement outcomes resulting from the use of this methodology would allow practitioners to adopt interventions validated as effective rather than denying such services to students because other standards of alternative approaches (e.g. RFT) have not been met. Bulterman-Bos (2008) suggested that researchers must adopt a view of education as moral practice. Her belief was that the way researchers conceptualise education dictates how they practise research and also how the results of the work they do is communicated. In this regard, even the results of well-designed RFT studies cannot be

universally applied with the expectation that such results will be transferable. Rather, all research, including RFT studies, requires extensive replication in order to determine the degree to which contextual factors potentially affecting fidelity of implementation result in divergent outcomes that cause practitioners to question the utility of the original research. In contrast, the use of multiple-baseline design logic conceptualises education as a system made up of multiple loosely-coupled systems (Weick, 1976) that are real world representations of the contexts in which large-scale reform occurs. As a result, any findings done across such diverse contexts are better able to immediately and accurately communicate relevant information to educational decision-makers.

Addressing the timeliness and utility of evaluative information

Recent federal, state, and local programs that place highly aggressive timelines associated with mandated student performance goals have emphasised the importance of timeliness. The problem raised in this paper is that the very RFT methodological approach which is intended to provide valid conclusions of individual studies is also overly restrictive insofar as being considered as the only source of timely information for educational leaders faced with decisions of selecting instructional alternatives for school reform (Henig, 2008/2009; Hess, 2008/2009; Ronka et al., 2008/2009). While well-conducted RFT studies can yield valid information estimates, subsequent replication of such studies is still a necessary design requirement to confirm or disconfirm research conclusions. It is the requirement that such large-scale, resource intensive studies be repeated and that in turn causes the considering of such studies as the sole source for establishing evidence-based effectiveness a questionable issue. Given the preceding, the need by practitioners and decision makers to address performance issues in a timely manner is hindered by the extensive RFT design requirements.

Emphasising the concept of replicability in evaluative research

An important concept, applicable to any form of empirical research initiative is that the replicability of findings is both a general scientific perspective and a necessary requirement for establishing research conclusions, and are better than findings from an individual study, no matter how well such an individual study is designed. In emphasising replicability, the logical requirements for multiple-baseline designs (see Sidman, 1960) provides a design framework for the aggregation of evaluative findings about the effectiveness of a particular instructional intervention that is obtained from diverse local school settings. In effect, such designs involve inter-study replication of the effects of an experimental intervention across what Stanley and Campbell (1963) have called “time series”.

As an example illustrating of multiple baseline design logic, an experimental study might involve three experimental units (e.g., schools) from which baseline data would be obtained for a series of intervals (e.g., tests by years), after which the experimental intervention would be introduced to one randomly-selected unit while data collection continued for all three. Then, after the experimental effect of the intervention stabilised,

the intervention would be continued with the first unit and implemented with a second randomly selected unit. Again, data collection would continue from all three sites until the effect of the second treatment stabilised as well.

The point of such an experimental study would be to demonstrate that the experimental effect observed (in comparison to the performance obtained without the treatment being implemented) would result after the intervention was implemented. And, the resulting inferential form of “causal” conclusion would be that the experimental effect could be produced by implementing the instructional treatment. Of course, if the expected effect did not occur, then the study would be classified a failure. But the important points from the standpoint of experimental design are (a) that the emphasis in the overall design is on the time-lagged replication of effects and (b) that the logical scope of the design subsumes that of randomised field trials (RFAs) for which the randomised assignment of treatment is not time-lagged (i.e., not implemented at the same point in time).

While logically powerful, implementing multiple-baseline experimental or quasi-experimental designs typically are not feasible in school evaluation settings because of the resource-intensive, multiyear implementation/instrumentation requirements. However, multiple-baseline design logic is readily applicable to evidence-based conclusions that result from the aggregation of school-based pre-post evaluative data when the instructional intervention of interest has been implemented in a time-lagged fashion across schools in different settings. The point of this paper is that, when appropriate existing forms of evaluative data that meet the multiple baseline design logic requirements can be obtained and analysed, the results of aggregating such data can provide informational support to decision makers forced to make decisions in “real time” regarding the potential effectiveness of an instructional intervention.

Using the logic of multiple-baseline designs as an evaluative model

The form of conclusion resulting from a multiple-baseline design is always in the form of an assertion that a specified effect was produced by implementing an intervention. This form of conclusion is consistent across all domains of scientific investigation that utilise experimental research methodology. The reasoning justifying the conclusion - based on successful and multiple replication of a study - is straightforward. It is simply, if the application of the intervention has consistently produced a desired effect in the past, then it would be expected to produce that effect again if implemented with fidelity across comparable settings. Certainly such an expectation meets the methodological requirement of testability (i.e., falsification). In fact, interpreted probabilistically, such reasoning is fundamental to the establishment of all scientific knowledge.

Given the preceding, the benefits to educational leaders of using such a multiple-baseline design framework are the following. Although all experimental studies require implementation of an intervention under control of the researcher(s), a pattern of data following the logic of a multiple-baseline design can yield meaningful conclusions as long as (a) comparable data can be used to integrate the effects of an intervention across

different (and independent) sites and (b) the adoption of the interventions are time-staggered. Such a rationale provides the means to integrate independent components that, without such integration, would not yield interpretable data. In effect, the argument advanced is: If an intervention can be shown to produce results on a before-after intervention context across many independent sites, then, from an evaluative standpoint, the resulting data can be interpreted as “causal”, subject to probabilistic qualifications.

In the following example, publically available publisher’s data were used to illustrate the use of multiple-baseline designs as an evaluative tool using existing pre-post intervention and post-intervention performance trends associated with results of the implementation of school reform reading initiatives across a wide range of schools in diverse contexts over time. While interpreting such findings raises questions regarding sampling bias (i.e., only positive findings are reported), the collection of findings across schools represents a correlational form of a multiple-baseline design with missing data that, as such, is amenable to multi-level statistical analysis. From the standpoint of replicability, the resulting form of statistical analysis provides a meaningful framework for the evaluation of any new specific intervention implemented in multiple schools sites.

An applied example illustrating application of the model

Presented in support of the argument is an illustrative set of evaluative data that does not meet RFT standards, but, with the supporting multiple-baseline rationale, allows potentially sound evaluative conclusions about the effectiveness of an intervention. The methodological issues presented argue for enhancing the usability of school-based evaluative outcome data of instructional programs by educational decision makers. The following shows how an approach using multiple-baseline design logic can provide alternative means of providing educational decision makers with evidence of instructional intervention effectiveness based on replicability of findings.

Participants

Participants were a total of 77 elementary and middle schools implementing direct instruction reading programs (see below) between 1996 and 2008 and reporting pre- and post- implementation performance of student reading performance by year as measured by state or nationally normed reading tests. The participating schools were demographically diverse (e.g., percent minority ranged from 3% to 100%, free/reduced lunch ranged from 0% to 99%).

Interventions and programs

The interventions consisted of the family of direct instruction reading programs. Primarily, Reading Mastery was implemented in grades K-5, while Corrective Reading was implemented in grades 3-8. Horizons, a program for grades K-3 representing a combination of instructional strategies and formats from Reading Mastery and Corrective Reading, also was used in several schools. This family of programs has an extensive history of evidence-based success across a variety of settings and contexts (Marchand-

Martella, Slocum, & Martella, 2003; Scientific Research Associates, n.d.). Direct instruction reading programs are characterised by utilising structured and systematic learning environments that emphasise (a) developmental sequencing of ideas and skills, (b) application based outcomes, (c) data driven teaching, and (d) frequent and targeted feedback. The programs are designed for students and adults that provide students with instruction ranging from remedial skills to advanced application oriented concepts.

Design

The purpose of the research design was to illustrate the potential informational value of a series of replicated studies for educational decision makers - even though the set as a whole suffered from selection bias. Overall, the design was a quasi-experimental, multiple-baseline design multilevel regression model that used data-by-year that always included prior- and post-intervention results. The implementation of the interventions across schools was distributed across a multi-year period that differed across schools. This characteristic of the data allowed for the use of the multilevel repeated measures design illustrating that a multi-site multi-year stratified implementation of a school reform can be successfully analysed using data readily available to school reform practitioners and researchers as is appropriate.

Data sources and multilevel analysis

All of the data were obtained from descriptive studies reported on the SRA web-site. Because the data were descriptive summaries (e.g., student achievement levels by year-prior to and after the intervention), typical meta-analytic hierarchical linear modeling (HLM) approaches incorporating effect-size (Raudenbush & Bryk, 2002) could not be used because error estimates were not available. However, as a practical substitute, data were scaled to means of zero and standard deviations of unity within schools (or grades within schools) in order to insure that pre/post implementation trends and post-implementation trends obtained from different reading achievement measures (e.g., state, norm-referenced) were comparable across schools.

In the 2-level HLM statistical model used, school demographic characteristics (percent of minority students, percent of students on free/reduced lunch) were coded at level 2 and the multi-year information nested within schools coded at level 1. In assigning the level 1 variables for the HLM model, a dummy variable for treatment indicated whether achievement data were obtained prior to or after the intervention (0= prior, 1= after) and served as a test for the pre/post intervention effect. Although the majority of schools reported achievement data by grade, in a few schools achievement data were reported for a group of grades (e.g., for grades 3-4-5). For those schools, the mean grade was used for the grade-level predictor. In addition, to assess the effects of the reading intervention after it was initiated, the number of years using the reading program was coded as a second treatment predictor (e.g., initial/first year of implementation = 0, second year = 2, etc.) and years prior to the reading intervention coded as -1, -2, etc.). In addition, to the two coded treatment variables, grade level and standardised within-school reading achievement were included as Level 1 variables in the HLM model.

Specifically the two level HLM model used was the following:

Level-1 Model

$$\text{Reading_Achievement}_{ij} = \beta_{0j} + \beta_{1j} * (\text{Test_1}_{ij}) + \beta_{2j} * (\text{Test_2}_{ij}) + \beta_{3j} * (\text{Grade_Level}_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{Pct_Min}_{ij}) + \gamma_{02} * (\text{Pct_FRL}_{ij}) + u_{0j};$$

$$\beta_{1j} = \gamma_{10} + u_{1j};$$

$$\beta_{2j} = \gamma_{20} + u_{2j};$$

$$\beta_{3j} = \gamma_{30} + u_{3j};$$

Where, for Level 1 variables: Test_1 = 1 if prior to the intervention, 0 if after; Test_2 = linear coefficients assigned to years after implementation; Grade_Level = Grade Level for the data (in a few cases, for schools having multiple grade levels, average grade was used); Reading_Achievement = standardised within school reading achievement.

For Level 2 variables: Pct_Min = percent of minority students in a school; Pct_FRL = percent of students in a school receiving free or reduced lunch. (Note: r_{ij} and $(u_{0j}, u_{1j}, u_{2j}, u_{3j})$ are Level 1 and Level 2 error terms, respectively.

The rationale for the HLM model used was to consider schools as if they were individuals for whom repeated measures were available for analysis (i.e., considering such repeated measures nested within individuals or, in the present application, as years within schools). In testing model components, HLM computes regression coefficients appropriate for variables at each level. In the case of the major treatment variable, the HLM coefficient for treatment (coded as 1 or 0) indicated the overall effect of the reading program. In addition, the HLM coefficient for the post-treatment treatment variable indicated whether the effect of the treatment accelerated after initial implementation.

Results

Overall, the multilevel analysis model used allowed each school to serve as its own control with regard to the time-lagged, pre- vs. post-intervention achievement differences. The analysis addressed the general question of whether the interventions resulted improved achievement after the intervention was introduced. Table 1 summarised the results for the full HLM6 model. As Table 1 shows, The HLM6 results found that the pre-post difference in standardised student reading achievement was statistically significant, $t(573) = 5.29$, $p < .001$, as was the linear acceleration in student achievement following the initial year of the intervention, $t(573) = 8.66$, $p < .001$. The percentage of students receiving free/reduced lunch, a general SES measure was also found significant, $t(71) = 2.25$, 5.29 , $p < .03$, but the percent of minority students in the school and grade level were not.

The statistical analysis of these school-based school-level evaluative data reported by SRA showed that the introduction of the direct instruction reading programs across a wide variety of school sites was followed by a significant improvement in student reading

Table 1: HLM analysis of the 2-level model

Fixed effect		Standard. approximation				
		Coefficient	Error	T-ratio	d.f.	P-value
For INTRCPT1, β_0	INTRCPT2, γ_{00}	-1.017	0.33	-3.05	71	0.00
	Pct_Min, γ_{01}	0.003	0.002	1.35	71	0.18
	Pct_FRL, γ_{02}	0.006	0.003	2.25	71	0.03
For Test_1 slope, β_1	INTRCPT2, γ_{10} a	0.570	0.108	5.29	573	0.00
For Test_2 slope, β_2	INTRCPT2, γ_{20} b	0.220	0.025	8.66	573	0.00
For Grade_Level slope, β_3	INTRCPT2, γ_{30}	-0.040	0.021	-1.87	573	0.06

a. 95% Confidence Interval for Test_1 Treatment is: [+0.36, +0.78]

b. 95% Confidence Interval for Test_2 is: [+0.17, +0.27]

achievement and that the achievement levels displayed by students increased with continued use of the programs and that these effects were consistent across grade levels. In interpreting these findings (see Table 1), educational decision makers could expect that the initial year of implementation of such programs would result in a pre-post implementation increase of .57 standard deviations of their within-school standardised reading achievement level (i.e., z equivalent = .57). In addition, as shown in Figure 1, beginning with year 2 of implementation, the .57 achievement expectation also would increase an additional .22 standardised units per year (i.e., Effect for Year 2 = .79; for Year 3 = 1.01; for Year 4 = 1.23).

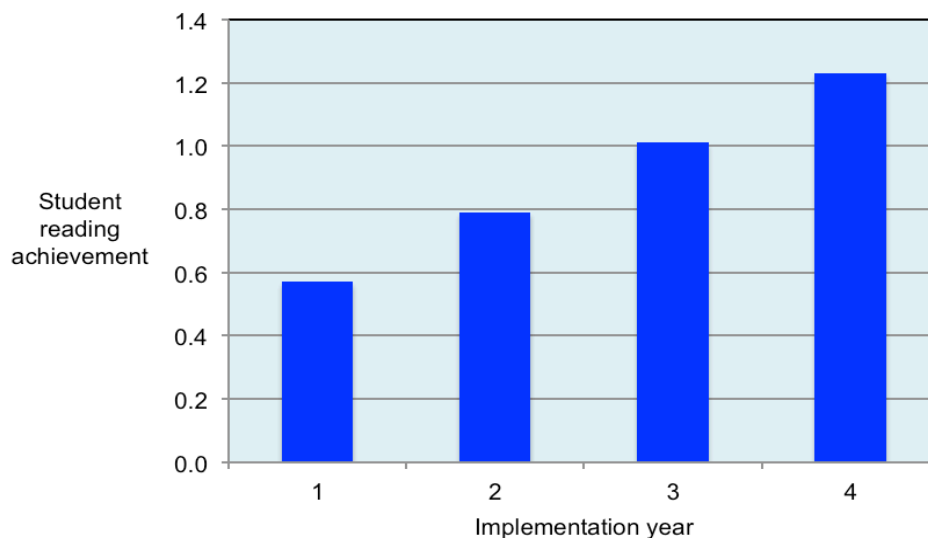


Figure 1: Effect of reading curriculum on reading achievement over a four year period

Since these data represented multi-year implementation periods, it is possible that the increased achievement trends reflected both improved teacher skills in program delivery as

well as the cumulative effect of the reading curriculum on student achievement levels as they entered succeeding grades. Because, consistent with a multiple-baseline design requirement, the initiation of the treatments was distributed across a multi-year time span, the group of findings reported across the multiple schools demonstrated effective replication of the effect of the programs (see also Adams & Engelmann, 1996).

Discussion

The use of RFT in education is an important and sound methodological approach for providing evidence-based conclusions of instructional interventions. However, it is not without serious limitations because of the extensive resources required to implement such studies. Because of such RFT limitations, alternative methodological approaches have become a topic for discussion in the literature (see Murnane & Willett, 2011). The application of multiple baseline design logic to interpret the findings resulting from the aggregation of existing evaluative data across multiple school sites is a specific alternative that offers educational decision makers methodologically-sound information that, by argument, are as valid as RFT studies. Given that findings from RFT studies are not widely available, considering them as the unique source of evidence-based information in the face of the present alternative is far too unnecessarily limiting in providing of research conclusions to educational leaders. With this point in mind, Kochanek and Clifford (2011) have argued that:

District administrators appear to place a higher value on information coming from other districts than that coming from the research community or state education agency. When discussing sources of information for these two decision points and the frequent use of other districts as models, respondents talked about the value they placed on using the work of those who have actually put ideas into practice. (p.13)

Additionally, the diverse sources of local data used in the methodological approach presented in this paper incorporate the points raised by Kochanek and Clifford (2011) regarding the connecting of the findings to practitioner decision makers.

The study example in this paper that illustrated the use of data that were generated as a result of a series of interventions in working school settings. A broad perspective is that with the emphasis on local school evaluation, the approach advocated in this paper may become more important in supporting school administrators in their decision making efforts than RFT. This perspective is consistent with the research on knowledge diffusion and context (see Rogers, 2003) and the idea of relevancy (Fusarelli, 2008). Additionally, Coburn et al (2008) posited that interpretation is one of three critical stages that educators transition through when using research. As shown here, multiple-baseline design logic can use existing data in a manner that is more readily interpretable because of how the data were generated and the ease with which educational leaders could, through follow-up contacts with other educators, estimate future results from adopting the intervention with fidelity.

The use of the multiple-baseline framework presented suggests the usefulness of aggregating disparate evaluative data into useful patterns focusing on replication of findings. Because of the emphasis in the design upon inter-site replication of time-distributed interventions, educational decision makers are able to consider such findings as providing a “causal conclusion” of program effectiveness that is far stronger than a “proof-of-concept” demonstration based on pre-post data alone. Clearly such findings provide an evidenced-based perspective to consider the feasibility of adopting such interventions that is far better than the “no information available” that would result from waiting from studies to be conducted that meet RFT methodological requirements. Given the importance of addressing educational needs, the emphasis on the replicability of evaluative findings when arranged in multiple-baseline design logic has the potential to be a useful form of evaluation information for such decision makers.

Conclusions and implications

This paper presented a scenario illustrating how multiple-baseline design logic could use data generated in real-world settings in a manner that has potential for use by school personnel as they engage in research and evaluate in research findings. Educational leadership has struggled to utilise evidence-based reforms for a number of reasons (Anneburg Institute for School Reform, 2005) including the restrictive use of some research designs, lack of capacity of educational leaders, weak or non-existing partnerships between researchers and schools, and an overall misapplication of research findings. While this paper does not offer solutions for all the barriers faced by researchers and school reformers as that attempt to develop evidence-based improvement it does speak to the some of the concerns educators raise with respect to applicability of findings, large scale implementation, and the use of existing data that requires fewer resources and restrictions than other approaches. The methodology presented in this study provides educational leaders evidence on school reform that is generated in settings advocated by Ellis (2005) by using large multi-site and multi-year contexts to address concerns expressed by educational leaders and other researchers on the diffusion of innovations (see Rogers, 2003).

The engagement in research by local educators has been argued (see Cooper, Levin, & Campbell, 2009; Honig, 2007; *Education Week*, 2013) as a basis for forming partnerships to promote the use of research in real school settings. The formation of such partnerships is important because it potentially provides the means through which local school districts could identify and pool existing pre-post implementation achievement data for analysis using the combination of multiple baseline design and multilevel analysis methodology presented in this paper. Further, through such partnerships, a database for the cumulative collection of such evaluative data could be established. Overall, such partnerships would provide the means for collaborative data collection, data analysis, and dissemination of findings. And, as part of such disseminations, educational leaders using findings would be able to pursue follow-up contact with other leaders in demographically similar school districts to identify implementation requirements. Overall, as pre-post achievement data from multiple content-similar school settings is obtained, the model presented here has promising implications for advancing sound, evidence-based school decision-making.

References

- Adams, G. & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISRAR*. Seattle, WA: Educational Achievement Systems.
- Burch, P. (2007). Educational policy and practice from the perspective of institutional theory: Crafting a wider lens. *Educational Researcher*, 36(2), 84-95.
<http://dx.doi.org/10.3102/0013189X07299792>
- Bulterman-Bos, J. (2008). *Response to comments: Clinical study: A pursuit of responsibility as the basis of educational research*. *Educational Researcher*, 37(7), 439-445.
<http://dx.doi.org/10.3102/0013189X08326296>
- Chatterji, M. (2008). *Comments on Slavin*: Synthesizing evidence from impact evaluations in education to inform action. *Educational Researcher*, 37(1), 23-26,
<http://dx.doi.org/10.3102/0013189X08314287>
- Coburn, C. E., Honig, M. & Stein, M. K. (2008). What is the evidence on districts' use of evidence? In L. Gomez, J. Bransford & D. Lam (Eds.), *Research and practice: The state of the field*. Cambridge, MA: Harvard Education Press.
- Cooper, A., Levin, B. & Campbell, C. (2009). The growing (but still limited) importance of evidence in education policy and practice. *The Journal of Educational Change*, 10(2-3), 159-171. <http://dx.doi.org/10.1007/s10833-009-9107-0>
- Dobbins, M., Rosenbaum, P., Plews, N., Law, M. & Fysh, A. (2007). Information transfer: What do decision makers want and need from researchers? *Implementation Science*, 2:20.
<http://www.implementationscience.com/content/2/1/20>
- Education Week* (2013). Spotlight: On data driven decision making.
<http://www.edweek.org/ew/marketplace/products/spotlight-data-driven-decisionmaking-v2.html?cmp=EB-SPT-032113>
- Ellis, A. (2005). *Research on educational innovations*. Larchmount, NY: Eye on Education.
- Flyvbjerg, B. (2001). *Making social science matter. Why social inquiry fails and how it can succeed again*. Cambridge, UK: Cambridge University Press.
- Fusarelli, L. (2008). Flying (partially) blind: School leaders' use of research in decision-making. *Phi Delta Kappan*, 89(5), 365-368.
<http://www.kappanmagazine.org/content/89/5/365.abstract>
- Henig, J. R. (2008/2009). The spectrum of educational research. *Educational Leadership*, 66(4), 6-11. <http://www.ascd.org/publications/educational-leadership/dec08/vol66/num04/The-Spectrum-of-Education-Research.aspx>
- Hess, F. M. (2008/2009). The new stupid. *Educational Leadership*, 66(4), 12-17.
<http://www.ascd.org/publications/educational-leadership/dec08/vol66/num04/The-New-Stupid.aspx>
- Hess, F. (2007). *When research matters*. Cambridge, MA: Harvard Education Press.
- Honig, M. I. & Coburn, C. E. (2007). Evidence-based decision-making in school district central offices: Toward a policy and research agenda. *Educational Policy*, 22(4), 578-608.
<http://dx.doi.org/10.1177/0895904807307067>
- IES (Institute of Education Sciences) (2013). Requests for applications. U.S. Department of Education. <http://ies.ed.gov/funding/14rfas.asp>

- Kochanek, J. & Clifford, M. (2011). Refining a theory of knowledge diffusion among district administrators. A paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA. (April 9, 2011)
- Lagemann, E. (2002). *Usable knowledge in education research*. New York: Spencer Foundation.
- Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *The Review of Higher Education*, 30(1), 1-19.
http://muse.jhu.edu/journals/review_of_higher_education/toc/rhe30.1.html
- Levin, B. (2010). Leadership for evidence-informed education. *School Leadership & Management*, 30(4), 303-315. <http://dx.doi.org/10.1080/13632434.2010.497483>
- Marchand-Martella, N. E., Slocum, T. A. & Martella, R. C. (2003). *Introduction to direct instruction*. Columbus, OH: Allyn & Bacon.
- Maynard, A. (2007). Translating evidence into practice: Why is it so difficult? *Public Money and Management*, 27(4), 251-256. <http://dx.doi.org/10.1111/j.1467-9302.2007.00591.x>
- Moss, B. & Yeaton, W. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, 28(3), 215-229. <http://dx.doi.org/10.3102/01623737028003215>
- No Child Left Behind Act of 2001* Pub. L. No. 107-110, 115 Stat. 1425 (2002).
<http://www.ed.gov/policy/elsec/leg/esea02/>
- Nutley, S., Walter, I. & Davies, H. T. O. (2007). *Using evidence*. Bristol: The Policy Press.
- Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, 52, 501-525.
<http://dx.doi.org/10.1146/annurev.psych.52.1.501>
- Raudenbush, S. W. & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Rickinson, M. (2005). *Practitioners' use of research: A research review for the National Evidence for Education Portal (NEEP) Development Group*. (Working Paper). London: National Educational Research Forum. <http://www.eep.ac.uk/nerf/word/WP7.5-PracuseofRe42d.doc?version=1>
- Ronka, D., Lachat, M. A., Slaughter, R. & Meltzer, J. (2008/2009). Answering the questions that count. *Educational Leadership*, 66(4), 18-24. <http://www.ascd.org/publications/educational-leadership/dec08/vol66/num04/Answering-the-Questions-That-Count.aspx>
- Rogers, E. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Scientific Research Associates (n.d.). *What is direct instruction?* https://www.mheonline.com/assets/sra_download/ReadingMasterySignatureEdition/MoreInfo/DI_Method_2008.pdf
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Slavin, R. E. (2008a). *Response to comments: Evidence-based reform in education: Which evidence counts?* *Educational Researcher*, 37(1), 47-50.
<http://dx.doi.org/10.3102/0013189X08315082>
- Slavin, R. E. (2008b). Perspectives on evidence-based reform in education - What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5-14. <http://dx.doi.org/10.3102/0013189X08314117>
- Sloane, F. (2008). *Comments on Slavin: Through the looking glass: Experiments, quasi-experiments, and the medical model*. *Educational Researcher*, 37(1), 41-46.
<http://dx.doi.org/10.3102/0013189X08314835>
- Stuart, E. A. (2007). Estimating causal effects using school-level data sets. *Educational Researcher*, 36(4), 187-198. <http://dx.doi.org/10.3102/0013189X07303396>

- Van der Heyden, A., Witt, J. & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, 45(2), 225-256.
<http://dx.doi.org/10.1016/j.jsp.2006.11.004>
- Weick, K. (1976). Educational organizations as loosely-coupled systems. *Administrative Science Quarterly*, 21(1), 1-9. <http://www.jstor.org/stable/2391875>
- West, S., Duan, N., Pequegant, W., Gasit, P., Des Jarlais, D., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M. & Mullen, P. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, 98(8), 1359-1366.
<http://dx.doi.org/10.2105/AJPH.2007.124446>
- William, D. & Lester, F. K. Jr. (2008). On the purpose of mathematics education research: Making productive contributions to policy and practice. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 32-48). New York: Routledge.

Dr Theodore S. Kaniuka is Associate Professor of Educational Leadership in the Department of Educational Leadership at Fayetteville State University. His present research interests are in the areas of high school reform, in particular Early College High Schools, research methods and program evaluation, and educational policy.
Email: tkaniuka@uncfsu.edu

Dr Michael R. Vitale is Professor of Educational Research in the Department of Curriculum and Instruction at East Carolina University. His research includes the development of models and operational instructional initiatives to raise school achievement expectations as well as instructional design principles to undergraduate teacher education programs.
Email: vitale@ecu.edu

Dr Nancy R. Romance is Professor of Science Education in the Department of Teaching and Learning at Florida Atlantic University. She holds degrees in Educational Leadership and in Science. Her research interests include a K-5 model that integrates literacy within in-depth science instruction and student vocabulary development.
Email: romance@fau.edu

Please cite as: Kaniuka, T. S., Vitale, M. R. & Romance, N. R. (2013). Aggregating school based findings to support decision making: Implications for educational leadership. *Issues in Educational Research*, 23(1), 69-82.
<http://www.iier.org.au/iier23/kaniuka.html>